

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 01:08:33 ; Search time 725.234 Seconds
(without alignments)
1536.704 Million cell updates/sec

Title: US-10-673-854-1

Perfect score: 23
Sequence: 1 catgtatttgatgggagagagg 23

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 23 | 100.0 | 68726 | 9 AC016681 | AC016681 Homo sapi |
| 2 | 23 | 100.0 | 100545 | 9 AL590379 | AL590379 Human DNA |
| 3 | 23 | 100.0 | 129357 | 9 AL590492 | AL590492 Human DNA |
| 4 | 19.8 | 86.1 | 285888 | 2 AC107131 | AC107131 Rattus no |
| 5 | 18.8 | 81.7 | 3191 | 10 BC070398 | BC070398 Mus muscu |
| 6 | 18.8 | 81.7 | 5634 | 9 AB023160 | AB023160 Homo sapi |
| 7 | 18.8 | 81.7 | 8225 | 10 MMU420922 | AJ420922 Mus muscu |
| 8 | 18.8 | 81.7 | 20838 | 9 AC117508 | AC117508 Homo sapi |
| 9 | 18.8 | 81.7 | 25901 | 9 AC139617 | AC139617 Homo sapi |
| 10 | 18.8 | 81.7 | 134296 | 9 HS352H24 | AL449215 Homo sapi |
| 11 | 18.8 | 81.7 | 151700 | 9 AC133528 | AC133528 Homo sapi |
| 12 | 18.8 | 81.7 | 163947 | 2 AC102466 | AC102466 Mus muscu |
| 13 | 18.8 | 81.7 | 167797 | 2 AC151841 | AC151841 Mus muscu |
| 14 | 18.8 | 81.7 | 171446 | 2 AC118744 | AC118744 Mus muscu |
| 15 | 18.8 | 81.7 | 175280 | 10 AC113270 | AC113270 Mus muscu |
| 16 | 18.8 | 81.7 | 176364 | 10 AC113126 | AC113126 Mus muscu |
| 17 | 18.8 | 81.7 | 185571 | 9 AC133781 | AC133781 Homo sapi |
| 18 | 18.8 | 81.7 | 191154 | 2 AC132150 | AC132150 Homo sapi |
| 19 | 18.8 | 81.7 | 193634 | 2 AC068495 | AC068495 Mus muscu |

| | | | | | | |
|------|------|------|--------|----|------------|--------------------|
| c 20 | 18.8 | 81.7 | 193894 | 9 | AC114730 | AC114730 Homo sapi |
| c 21 | 18.8 | 81.7 | 200711 | 2 | AC133960 | AC133960 Homo sapi |
| c 22 | 18.8 | 81.7 | 229639 | 5 | AC140947 | AC140947 Gallus ga |
| c 23 | 18.8 | 81.7 | 238704 | 2 | EX890628 | EX890628 Mus muscu |
| c 24 | 18.8 | 81.7 | 258520 | 10 | AL662812 | AL662812 Mouse DNA |
| c 25 | 18.4 | 80.0 | 450 | 8 | ATSRG3PRT | X98376 A.thaliana |
| 26 | 18.4 | 80.0 | 1117 | 8 | AY114073 | AY114073 Arabidops |
| 27 | 18.4 | 80.0 | 1292 | 8 | AY089160 | AY089160 Arabidops |
| 28 | 18.4 | 80.0 | 1307 | 8 | AY072127 | AY072127 Arabidops |
| 29 | 18.4 | 80.0 | 1501 | 8 | BT003327 | BT003327 Arabidops |
| 30 | 18.4 | 80.0 | 3050 | 6 | C0849723 | C0849723 Sequence |
| 31 | 18.4 | 80.0 | 3050 | 9 | AK126763 | AK126763 Homo sapi |
| 32 | 18.4 | 80.0 | 74188 | 2 | AC101144 | AC101144 Mus muscu |
| 33 | 18.4 | 80.0 | 103960 | 8 | ATAC011664 | AC011664 Arabidops |
| c 34 | 18.4 | 80.0 | 110000 | 2 | AC091341 | AC091341 Rattus no |
| c 35 | 18.4 | 80.0 | 110000 | 2 | AC114711_3 | Continuation (4 of |
| c 36 | 18.4 | 80.0 | 158599 | 9 | AC108066 | AC108066 Homo sapi |
| 37 | 18.4 | 80.0 | 163277 | 2 | AC074259 | AC074259 Trypanoso |
| c 38 | 18.4 | 80.0 | 163463 | 10 | AC134536 | AC134536 Mus muscu |
| c 39 | 18.4 | 80.0 | 173268 | 2 | AC110225 | AC110225 Mus muscu |
| c 40 | 18.4 | 80.0 | 181683 | 10 | AC122180 | AC122180 Mus muscu |
| c 41 | 18.4 | 80.0 | 221631 | 9 | AC010867 | AC010867 Homo sapi |
| c 42 | 18.4 | 80.0 | 223728 | 2 | AC135443 | AC135443 Rattus no |
| c 43 | 18.4 | 80.0 | 240364 | 2 | AC114875 | AC114875 Rattus no |
| c 44 | 18.4 | 80.0 | 256385 | 2 | AC128721 | AC128721 Rattus no |
| c 45 | 18.4 | 80.0 | 306803 | 1 | AE017161 | AE017161 Prochloro |

ALIGNMENTS

| | | | | | |
|------------|---|------------|--------------------|--------|-----------------|
| RESULT 1 | AC016681 | 68726 bp | DNA | linear | PRI 30-SEP-2000 |
| LOCUS | Homo sapiens BAC clone RP11-62H15 | from Y, | complete sequence. | | |
| DEFINITION | AC016681 | | | | |
| ACCESSION | AC016681.2 | GI:7321924 | | | |
| VERSION | HTG. | | | | |
| KEYWORDS | Homo sapiens (human) | | | | |
| SOURCE | Homo sapiens | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 68726) | | | | |
| AUTHORS | Sulston,J.E. and Waterston,R. | | | | |
| TITLE | Toward a complete human genome sequence | | | | |
| JOURNAL | Genome Res. 8 (11), 1097-1108 (1998) | | | | |
| MEDLINE | 99063792 | | | | |
| PUBMED | 9847074 | | | | |
| REFERENCE | 2 (bases 1 to 68726) | | | | |
| AUTHORS | Joshu,C., Stromatt,C. and Wedgeworth,P. | | | | |
| TITLE | The sequence of Homo sapiens BAC clone RP11-62H15 | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 3 (bases 1 to 68726) | | | | |
| AUTHORS | Waterston,R.H. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (04-DEC-1999) Genome Sequencing Center, Washington | | | | |
| REFERENCE | 4 (bases 1 to 68726) | | | | |
| AUTHORS | Waterston,R.H. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (24-MAR-2000) Genome Sequencing Center, Washington | | | | |
| REFERENCE | 5 (bases 1 to 68726) | | | | |
| AUTHORS | Waterston,R.H. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (04-APR-2000) Genome Sequencing Center, Washington | | | | |
| REFERENCE | 6 (bases 1 to 68726) | | | | |
| AUTHORS | Waterston,R.H. | | | | |
| TITLE | Direct Submission | | | | |

```

JOURNAL Submitted (07-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 7 (bases 1 to 68726)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Mar 24, 2000 this sequence version replaced gi:6524399.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The position of this clone was established as part of a
collaboration between the Human Chromosome Y Mapping Project
(Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen,
and David C. Page at the Whitehead Institute for Biomedical
Research, Cambridge MA) and the Washington University Genome
Sequencing Center, St. Louis MO.

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-362J16; the clone sequenced
to the right is RP11-218E11. Actual start of this clone is at base
position 1 of RP11-62H15.
FEATURES             Location/Qualifiers
     source           1..68726
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="Y"
                     /map="Y"
                     /clone="RP11-62H15"
                     /clone_lib="RPCI-11"
     repeat_region    1..352
                     /rpt_family="L1"
     repeat_region    357..651
                     /rpt_family="Alu"
     repeat_region    676..769
                     /rpt_family="L1"
     repeat_region    1837..1873
                     /rpt_family="AT_rich"
     repeat_region    2274..2301
                     /rpt_family="AT_rich"
     repeat_region    2542..2572
                     /rpt_family="L1"

repeat_region       3768..3973      /rpt_family="AT_rich"
repeat_region       4258..4660      /rpt_family="MaLR"
repeat_region       5234..5568      /rpt_family="MaLR"
repeat_region       6159..6713      /rpt_family="MER1_type"
repeat_region       6718..6943      /rpt_family="L1"
repeat_region       6929..7694      /rpt_family="L1"
repeat_region       7709..7887      /rpt_family="L1"
repeat_region       7903..8247      /rpt_family="TA)n"
repeat_region       8264..9031      /rpt_family="L1"
repeat_region       9435..9583      /rpt_family="L1"
repeat_region       9584..9738      /rpt_family="MIR"
repeat_region       10428..10482     /rpt_family="MIR"
repeat_region       10574..11396     /rpt_family="L2"
repeat_region       11997..12024     /rpt_family="L1"
repeat_region       12040..13387     /rpt_family="AT_rich"
repeat_region       13390..13846     /rpt_family="L1"
repeat_region       13837..14070     /rpt_family="L1"
repeat_region       14074..15649     /rpt_family="L1"
repeat_region       15648..16717     /rpt_family="L1"
repeat_region       16720..16789     /rpt_family="L1"
repeat_region       16827..17303     /rpt_family="A-rich"
repeat_region       17417..17575     /rpt_family="L1"
repeat_region       17725..17768     /rpt_family="L2"
repeat_region       18047..18132     /rpt_family="AT_rich"
repeat_region       18338..18648     /rpt_family="L1"
repeat_region       19586..19620     /rpt_family="Alu"
repeat_region       19704..19918     /rpt_family="(CA)n"
repeat_region       19925..19980     /rpt_family="MIR"
repeat_region       20305..20330     /rpt_family="(ATG)n"
repeat_region       20333..20671     /rpt_family="(CA)n"
repeat_region       21471..21557     /rpt_family="MER2_type"
repeat_region       21661..22655     /rpt_family="MaLR"
repeat_region       22725..22971     /rpt_family="L1"
repeat_region       23441..23462     /rpt_family="L1"
repeat_region       23982..24011     /rpt_family="AT_rich"
repeat_region       24167..24258     /rpt_family="AT_rich"
repeat_region       /rpt_family="L1"

```

```

repeat_region 24399..24463
                /rpt_family="AT_rich"
repeat_region 24468..24576
                /rpt_family="L2"
repeat_region 24754..25021
                /rpt_family="Retroviral"
repeat_region 25115..25575
                /rpt_family="L2"
repeat_region 25715..26221
                /rpt_family="MaLR"
repeat_region 26501..26523
                /rpt_family="AT_rich"
repeat_region 26567..26925
                /rpt_family="MaLR"

```

```

Query Match      100.0%; Score 23; DB 9; Length 68726;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CATGTATTGATGGGGATAGAGG 23
Db 56609 CATGTATTGATGGGGATAGAGG 56631

```

```

RESULT 2
AL590379
LOCUS          AL590379 100545 bp DNA linear PRI 01-APR-2003
DEFINITION    Human DNA sequence from clone RP11-343H6 on chromosome X, complete
               sequence.
ACCESSION     AL590379
VERSION       AL590379.7 GI:29466482
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 100545)
AUTHORS       Howden, P.
TITLE         Direct Submission
JOURNAL       Submitted (01-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
               Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
               humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
               On Apr 1, 2003 this sequence version replaced gi:28933281.
COMMENT       ----- Genome Center
               Center: Wellcome Trust Sanger Institute
               Center code: SC
               Web site: http://www.sanger.ac.uk
               Contact: humquery@sanger.ac.uk
               -----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX> RP11-343H6 is from the library RPCI-11.2 constructed by the group

of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

```

FEATURES             Location/Qualifiers
     source           1..100545
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-343H6"
                     /clone_lib="RPCI-11.2"

```

ORIGIN

```

Query Match      100.0%; Score 23; DB 9; Length 100545;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CATGTATTGATGGGGATAGAGG 23
Db 67470 CATGTATTGATGGGGATAGAGG 67492

```

```

RESULT 3
AL590492
LOCUS          AL590492 129357 bp DNA linear PRI 11-DEC-2001
DEFINITION    Human DNA sequence from clone RP11-88H5 on chromosome X, complete
               sequence.
ACCESSION     AL590492
VERSION       AL590492.7 GI:15072594
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 129357)
AUTHORS       Heath, P.
TITLE         Direct Submission
JOURNAL       Submitted (10-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
               Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
               humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
               On Aug 1, 2001 this sequence version replaced gi:1329489.
               During sequence assembly data is compared from overlapping clones.
               Where differences are found these are annotated as variations
               together with a note of the overlapping clone name. Note that the
               variation annotation may not be found in the sequence submission
               corresponding to the overlapping clone, as we submit sequences with
               only a small overlap as described above.
               This sequence was finished as follows unless otherwise noted: all
               regions were either double-stranded or sequenced with an alternate
               chemistry or covered by high quality data (i.e., phred quality >=
               30); an attempt was made to resolve all sequencing problems, such
               as compressions and repeats; all regions were covered by at least
               one plasmid subclone or more than one M13 subclone; and the
               assembly was confirmed by restriction digest. The following
               abbreviations are used to associate primary accession numbers given
               in the feature table with their source databases: Em., EMBL; Sw.,
               SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
               database can be found at
               http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
               was generated from part of bacterial clone contigs of human
               chromosome X, constructed by the Sanger Centre Chromosome X Mapping
               Group. Further information can be found at
               http://www.sanger.ac.uk/HGP/ChrX
               RP11-88H5 is from the library RPCI-11.1 constructed by the group of
               Pieter de Jong. For further details see
               http://www.chori.org/bacpac/home.htm
               VECTOR: pBACE3.6

```

COMMENT

IMPORTANT: This sequence is not the entire insert of clone RP11-88H5 it may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone RP11-88H5 is at 1 in this sequence. The true left end of clone RP11-156J23 is at 129258 in this sequence.

Location/Qualifiers

FEATURES

```
source
1. .129357
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-88H5"
/clone_lib="RPC1-11.1"
2. .1884
/note="L1PB2 repeat: matches 4280. .6155 of consensus"
repeat_region
1933. .2582
/note="L1M4 repeat: matches 2318. .3033 of consensus"
repeat_region
misc_feature
complement(2286. .2752)
/note="match: GSS: Em:AQ450883"
misc_feature
complement(2291. .2748)
/note="match: GSS: Em:AQ360394"
misc_feature
2846. .3095
/note="match: STS: Em:HSC28P10"
repeat_region
2903. .3363
/note="MLT1H repeat: matches 5. .546 of consensus"
misc_feature
4611. .5105
/note="match: GSS: Em:AQ827938"
repeat_region
5147. .5317
/note="L1P4 repeat: matches 5423. .5594 of consensus"
repeat_region
5634. .6010
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region
6017. .6215
/note="MTC repeat: matches 1. .187 of consensus"
repeat_region
6369. .6505
/note="MTRC repeat: matches 254. .394 of consensus"
repeat_region
7390. .7767
/note="L1ME2 repeat: matches 4453. .4836 of consensus"
repeat_region
7776. .8020
/note="L1MC2 repeat: matches 6070. .6329 of consensus"
repeat_region
8028. .8148
/note="MER74A repeat: matches 445. .558 of consensus"
repeat_region
8174. .8227
/note="MER74A repeat: matches 412. .466 of consensus"
repeat_region
8228. .8523
/note="AluY repeat: matches 3. .297 of consensus"
repeat_region
8524. .8921
/note="MER74A repeat: matches 26. .412 of consensus"
repeat_region
8937. .9122
/note="L1MC1 repeat: matches 5870. .6072 of consensus"
repeat_region
9142. .9431
/note="AluY repeat: matches 1. .289 of consensus"
repeat_region
9432. .9818
/note="L1ME2 repeat: matches 4845. .5248 of consensus"
repeat_region
10083. .10417
/note="MLT1E repeat: matches 240. .568 of consensus"
repeat_region
10418. .10564
/note="L1ME2 repeat: matches 5237. .5397 of consensus"
repeat_region
10568. .11070
/note="MLT1D repeat: matches 1. .500 of consensus"
repeat_region
11071. .11816
/note="L1ME2 repeat: matches 5391. .6153 of consensus"
repeat_region
11825. .12014
/note="L1ME2 repeat: matches 1856. .2055 of consensus"
repeat_region
12146. .12258
/note="MLT1E repeat: matches 28. .136 of consensus"
repeat_region
12295. .12528
/note="MLT1E-internal repeat: matches 1114. .1335 of consensus"
repeat_region
12547. .12965
/note="MLT2CB repeat: matches 1. .433 of consensus"
repeat_region
13623. .14146
/note="MER9 repeat: matches 1. .511 of consensus"
repeat_region
14225. .14517
/note="MLT1-INTERNAL repeat: matches 291. .591 of consensus"
repeat_region
16863. .17118
/note="L1ME3A repeat: matches 5386. .5638 of consensus"
repeat_region
17119. .17250
/note="L1PB1 repeat: matches 6024. .6155 of consensus"
repeat_region
17251. .17506
/note="L1ME3A repeat: matches 5638. .5903 of consensus"
repeat_region
17546. .17932
/note="L1M4 repeat: matches 251. .627 of consensus"
repeat_region
18147. .18210
/note="L1M4 repeat: matches 616. .674 of consensus"
repeat_region
18184. .19717
/note="L1M4 repeat: matches 926. .2109 of consensus"
repeat_region
19782. .19970
/note="L1M4 repeat: matches 2123. .2308 of consensus"
repeat_region
19975. .20288
/note="AluO repeat: matches 12. .312 of consensus"
repeat_region
20314. .20637
/note="L1M4 repeat: matches 2316. .2663 of consensus"
repeat_region
20644. .21468
/note="L1P3 repeat: matches 15. .650 of consensus"
repeat_region
21469. .23040
/note="L1PA13 repeat: matches 4584. .6152 of consensus"
misc_feature
complement(23279. .23643)
/note="match: GSS: Em:B79974"
repeat_region
23281. .23580
/note="AluO repeat: matches 1. .291 of consensus"
repeat_region
23909. .24217
/note="AluX repeat: matches 1. .309 of consensus"
repeat_region
24276. .24322
/note="WADL1 repeat: matches 33. .79 of consensus"
repeat_region
24326. .24482
/note="L2 repeat: matches 1992. .2147 of consensus"
repeat_region
24990. .25140
/note="L1M4 repeat: matches 2404. .2560 of consensus"
repeat_region
25585. .25983
/note="L1 repeat: matches 3095. .3404 of consensus"
repeat_region
25903. .26200
/note="AluY5 repeat: matches 1. .296 of consensus"
repeat_region
26529. .26685
/note="FRAM repeat: matches -4. .152 of consensus"
repeat_region
26825. .26996
/note="MLT1J repeat: matches 320. .509 of consensus"
repeat_region
27032. .27116
/note="HERV16 repeat: matches 2. .89 of consensus"
repeat_region
27137. .27444
/note="LTR16A repeat: matches 115. .426 of consensus"
repeat_region
28310. .28924
/note="L1M4 repeat: matches 5064. .5692 of consensus"
repeat_region
29682. .29969
/note="AluX repeat: matches 1. .288 of consensus"
repeat_region
30022. .30136
/note="L1ME1 repeat: matches 6033. .6148 of consensus"
repeat_region
30157. .34776
/note="L1PA13 repeat: matches 648. .5299 of consensus"
repeat_region
34850. .35692
/note="L1PA13 repeat: matches -651. .215 of consensus"
repeat_region
35729. .36642
/note="L1 repeat: matches 3630. .4551 of consensus"
repeat_region
36652. .36809
/note="MLT1A1 repeat: matches 1. .160 of consensus"
repeat_region
36810. .38853
/note="L1P repeat: matches 3236. .5271 of consensus"
repeat_region
38877. .39968
/note="L1P repeat: matches 2318. .3413 of consensus"
repeat_region
40002. .40330
/note="L1P repeat: matches 2201. .2320 of consensus"
repeat_region
40118. .40327
/note="AluY repeat: matches 94. .303 of consensus"
repeat_region
40328. .41566
/note="L1P repeat: matches 2656. .3894 of consensus"
repeat_region
41571. .41865
/note="AluB repeat: matches 1. .288 of consensus"
repeat_region
41890. .41983
/note="L1 repeat: matches 3856. .3955 of consensus"
repeat_region
44982. .45187
/note="MLT1 repeat: matches 1. .205 of consensus"
repeat_region
45472. .45874
```



```

/note="MLr11 repeat: matches 1. .402 of consensus"
46448. .46782
/note="MER58B repeat: matches 2. .341 of consensus"
47373. .46157
/note="L1 repeat: matches 4164. .4992 of consensus"

Query Match      100.0%; Score 23; DB 9; Length 129357;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGTATTGATGGGATAGG 23
Db 97822 CATGTATTGATGGGATAGG 97844
|||||
|||||

RESULT 4
AC107131
LOCUS
DEFINITION
Rattus norvegicus clone CH230-38C17, *** SEQUENCING IN PROGRESS
***, 6 unordered pieces.
AC107131
VERSION
AC107131.4 GI:23267848
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 285888)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blawie,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgiou,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howell,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuehwa,L., Mahindartne,M., Mahmood,M., Malloy,K., Mangum,A.,
Maheshwari,M., Mahindartne,M., Lozado,R.J., Lu,X., Ma,J.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J.,
Steinle,T., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,P., Zhang,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 285888)
Worley,K.C.
Direct Submission
Submitted (16-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 285888)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 22, 2002 this sequence version replaced gi:21737099.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Alas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GONW
Center clone name: CH230-38C17
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 239711 bases at least Q40
Consensus quality: 243503 bases at least Q30
Consensus quality: 245994 bases at least Q20
Estimated insert size: 277420; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 11520: contig of 11520 bp in length
* 11521 11620: gap of unknown length
* 11621 244182: contig of 232562 bp in length
* 244183 244282: gap of unknown length
* 244283 247660: contig of 3378 bp in length
* 247661 247761: gap of unknown length
* 247761 264056: contig of 16296 bp in length
* 264057 264156: gap of unknown length
* 264157 268805: contig of 4649 bp in length
* 268806 268905: gap of unknown length
* 268906 285888: contig of 16983 bp in length.
* Location/Qualifiers
* .285888
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-38C17"
* 1399. .2262
* /note="clone_boundary"

```

misc_feature


```

RESULT 6
LOCUS      AB023160
DEFINITION Homo sapiens mRNA for KIAA0943 protein, partial cds.
ACCESSION AB023160
VERSION    AB023160.1
KEYWORDS   GI:4589529
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (sites)
AUTHORS    Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirotsawa,M.,
           Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE      Prediction of the coding sequences of unidentified human genes.
           XIII. The complete sequences of 100 new cDNA clones from brain
           which code for large proteins in vitro
JOURNAL    DNA Res. 6 (1), 63-70 (1999)
MEDLINE    99246063
PUBMED     10231032
REFERENCE  2 (bases 1 to 5634)
AUTHORS    Ohara,O., Nagase,T. and Kikuno,R.
TITLE      Direct Submission
JOURNAL    Submitted (04-FEB-1999) Oasamu Ohara, Kazusa DNA Research Institute,
           Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
           292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913,
           Fax: +81-438-52-3914)
FEATURES   Location/Qualifiers
            source
            1..5634
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="hh08433"
            /sex="male"
            /tissue="brain"
            /clone_lib="pBluescriptII SK plus"
            /dev_stage="adult"
            1..5634
            /gene="KIAA0943"
            <1..1107
            /gene="KIAA0943"
            /codon_start=1
            /product="KIAA0943 protein"
            /protein_id="BAA76787.1"
            /db_xref="GI:4589530"
            /translation="GATPLGSGVGRGKMDAAILTYDTLRFAPFDPFPETSEPVWILG
            RKYSIFTEKDEILSDVASRLWFTYRNFFPAIGTGTPTSDTGWGCMRCGQMIFQAQLV
            CRHLGRDWRWTQRKQSDSYFSLNAPIDRKDSYYSIHQIAQMGVGGKSGIQMGYGN
            TVAQVLKLAVFDTWSLAVHIAINDTVNMEIRLCRTSPVCAGATAPADSDRHCN
            GFPAGAEVTRPSPWRPLVLLIPLRLGLTDINBAAYVETLKHCFMPPQSLGVIGKENS
            AHYFIGVGBELIVLDPHTTQPAVEPTDGCIFIDESFHCQHPFCRMSIAELDPSIAVV
            RGHLSTQAFGACCLGMTRKTFTGFLRFFPSMLG"
ORIGIN
Query Match      81.7%; Score 18.8; DB 9; Length 5634;
Best Local Similarity 90.9%; Pred. No. 3.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2  ATGTATTGATGGGATAGAGG 23
      |||
Db      1556 ATATATTGATGGGAAAGAGG 1577

RESULT 7
LOCUS      MMU420922/c
DEFINITION Mus musculus partial PPARb/d gene for peroxisome proliferator
           activated receptor beta/delta, exons 4-8.
ACCESSION  AJ420922
VERSION    AJ420922.1
KEYWORDS   GI:22450050
           peroxisome proliferator activated receptor beta delta; PPARb/d
           gene.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE  1
AUTHORS    Larsen,L.K., Anri,E.Z., Mandrup,S., Pacot,C. and Kristiansen,K.
TITLE      Genomic organization of the mouse peroxisome proliferator-activated
           receptor beta/delta gene: alternative promoter usage and splicing
           yield transcripts exhibiting differential translational efficiency
JOURNAL    Biochem. J. 366 (Pt 3), 767-775 (2002)
MEDLINE    12051784
PUBMED     12059785
REFERENCE  2 (bases 1 to 8225)
AUTHORS    Larsen,L.K.
TITLE      Direct Submission
JOURNAL    Submitted (12-SEP-2001) Larsen L.K., Basic, Center for clinical and
           basic research, Ballerup Byvej 222, DK-2750 Ballerup, DENMARK
           Related sequences: AJ420918 - AJ420922.
FEATURES   Location/Qualifiers
            source
            1..8225
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /strain="C57 BL/6J"
            /db_xref="taxon:10090"
            1..6291
            /gene="PPARB/d"
            <1..181
            /gene="PPARB/d"
            /number=3
            182..336
            /gene="PPARB/d"
            /number=4
            /evidence=experimental
            /usedin=AJ420918:PPARBd SUTR_1A
            /usedin=AJ420918:PPARBd_mRNA_1A
            /usedin=AJ420919:PPARBd SUTR_1B1
            /usedin=AJ420919:PPARBd SUTR_1B2
            /usedin=AJ420919:PPARBd SUTR_1B3
            /usedin=AJ420919:PPARBd SUTR_1B4
            /usedin=AJ420919:PPARBd SUTR_1C
            /usedin=AJ420919:PPARBd_mRNA_1B1
            /usedin=AJ420919:PPARBd_mRNA_1B2
            /usedin=AJ420919:PPARBd_mRNA_1B3
            /usedin=AJ420919:PPARBd_mRNA_1B4
            /usedin=AJ420920:PPARBd_mRNA_1C
            /usedin=AJ420920:PPARBd SUTR_1D1
            /usedin=AJ420920:PPARBd SUTR_1D2
            /usedin=AJ420920:PPARBd SUTR_1D3
            /usedin=AJ420920:PPARBd SUTR_1D4
            /usedin=AJ420920:PPARBd_mRNA_1D1
            /usedin=AJ420920:PPARBd_mRNA_1D2
            /usedin=AJ420920:PPARBd_mRNA_1D3
            /usedin=AJ420920:PPARBd_mRNA_1D4
            /usedin=AJ420921:PPARBd_mRNA_CDS
            337..1927
            /gene="PPARB/d"
            /number=4
            1928..2066
            /gene="PPARB/d"
            /number=5
            /evidence=experimental
            /usedin=AJ420918:PPARBd SUTR_1A
            /usedin=AJ420918:PPARBd_mRNA_1A
            /usedin=AJ420919:PPARBd SUTR_1B1
            /usedin=AJ420919:PPARBd SUTR_1B2
            /usedin=AJ420919:PPARBd SUTR_1B3
            /usedin=AJ420919:PPARBd SUTR_1B4
            /usedin=AJ420919:PPARBd SUTR_1C
            /usedin=AJ420919:PPARBd_mRNA_1B1
            /usedin=AJ420919:PPARBd_mRNA_1B2
            /usedin=AJ420919:PPARBd_mRNA_1B3
            /usedin=AJ420919:PPARBd_mRNA_1B4
            /usedin=AJ420919:PPARBd_mRNA_1C

```

```

/usedin=AJ420920:PPARbd_5UTR_1D1
/usedin=AJ420920:PPARbd_5UTR_1D2
/usedin=AJ420920:PPARbd_5UTR_1D3
/usedin=AJ420920:PPARbd_5UTR_1D4
/usedin=AJ420920:PPARbd_mRNA_1D1
/usedin=AJ420920:PPARbd_mRNA_1D2
/usedin=AJ420920:PPARbd_mRNA_1D3
/usedin=AJ420920:PPARbd_mRNA_1D4
/usedin=AJ420921:PPARbd_CDS
2067. .3118
/genes="PPARb/d"
/number=5
3119. .3321
/genes="PPARb/d"
/number=6
/evidence=experimental
/usedin=AJ420918:PPARbd_5UTR_1A
/usedin=AJ420918:PPARbd_mRNA_1A
/usedin=AJ420919:PPARbd_5UTR_1B1
/usedin=AJ420919:PPARbd_5UTR_1B2
/usedin=AJ420919:PPARbd_5UTR_1B3
/usedin=AJ420919:PPARbd_5UTR_1B4
/usedin=AJ420919:PPARbd_5UTR_1C
/usedin=AJ420919:PPARbd_mRNA_1B1
/usedin=AJ420919:PPARbd_mRNA_1B2
/usedin=AJ420919:PPARbd_mRNA_1B3
/usedin=AJ420919:PPARbd_mRNA_1B4
/usedin=AJ420919:PPARbd_mRNA_1C
/usedin=AJ420920:PPARbd_5UTR_1D1
/usedin=AJ420920:PPARbd_5UTR_1D2
/usedin=AJ420920:PPARbd_5UTR_1D3
/usedin=AJ420920:PPARbd_5UTR_1D4
/usedin=AJ420920:PPARbd_mRNA_1D1
/usedin=AJ420920:PPARbd_mRNA_1D2
/usedin=AJ420920:PPARbd_mRNA_1D3
/usedin=AJ420921:PPARbd_mRNA_1D4
/usedin=AJ420921:PPARbd_CDS
3322. .3425
/genes="PPARb/d"
/number=6
3426. .3876
/genes="PPARb/d"
/number=7
/evidence=experimental
/usedin=AJ420918:PPARbd_5UTR_1A
/usedin=AJ420918:PPARbd_mRNA_1A
/usedin=AJ420919:PPARbd_5UTR_1B1
/usedin=AJ420919:PPARbd_5UTR_1B2
/usedin=AJ420919:PPARbd_5UTR_1B3
/usedin=AJ420919:PPARbd_5UTR_1B4
/usedin=AJ420919:PPARbd_5UTR_1C
/usedin=AJ420919:PPARbd_mRNA_1B1
/usedin=AJ420919:PPARbd_mRNA_1B2
/usedin=AJ420919:PPARbd_mRNA_1B3
/usedin=AJ420919:PPARbd_mRNA_1B4
/usedin=AJ420919:PPARbd_mRNA_1C
/usedin=AJ420920:PPARbd_5UTR_1D1
/usedin=AJ420920:PPARbd_5UTR_1D2
/usedin=AJ420920:PPARbd_5UTR_1D3
/usedin=AJ420920:PPARbd_5UTR_1D4
/usedin=AJ420920:PPARbd_mRNA_1D1
/usedin=AJ420920:PPARbd_mRNA_1D2
/usedin=AJ420920:PPARbd_mRNA_1D3
/usedin=AJ420921:PPARbd_mRNA_1D4
/usedin=AJ420921:PPARbd_CDS
3877. .4443
/genes="PPARb/d"
/number=7
4444. .>6291
/genes="PPARb/d"
/number=8
/evidence=experimental
/usedin=AJ420918:PPARbd_5UTR_1A

```

```

/usedin=AJ420918:PPARbd_mRNA_1A
/usedin=AJ420919:PPARbd_5UTR_1B1
/usedin=AJ420919:PPARbd_5UTR_1B2
/usedin=AJ420919:PPARbd_5UTR_1B3
/usedin=AJ420919:PPARbd_5UTR_1B4
/usedin=AJ420919:PPARbd_5UTR_1C
/usedin=AJ420919:PPARbd_mRNA_1B1
/usedin=AJ420919:PPARbd_mRNA_1B2
/usedin=AJ420919:PPARbd_mRNA_1B3
/usedin=AJ420919:PPARbd_mRNA_1B4
/usedin=AJ420919:PPARbd_mRNA_1C
/usedin=AJ420920:PPARbd_5UTR_1D1
/usedin=AJ420920:PPARbd_5UTR_1D2
/usedin=AJ420920:PPARbd_5UTR_1D3
/usedin=AJ420920:PPARbd_mRNA_1D4
/usedin=AJ420921:PPARbd_CDS
6286. .6290
/genes="PPARb/d"

polyA_signal
ORIGIN
      Query Match      81.7%; Score 18.8; DB 10; Length 8225;
      Best Local Similarity 90.9%; Pred. No. 3.1e+02;
      Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CATGTATTGATGGGGATAGAG 22
      ||||| ||||| ||||| |||||
Db      5116 CATGCTTTGAGGGGATAGAG 5095

RESULT 8
AC117508
LOCUS   Homo sapiens 3 BAC RP11-207P12 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC117508
VERSION   AC117508.5 GI:28882126
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.K., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,I.C.,
Lewis,L., Li,O., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,

```



```

repeat_region      /rpt_family="LTR26B"
                    complement(9228..8482)
repeat_region      /rpt_family="AluJo"
                    complement(9483..8994)
repeat_region      /rpt_family="LTR26B"
                    9177..9230
repeat_region      /rpt_family="L2"
                    complement(9907..9951)
repeat_region      /rpt_family="MIRb"
                    10017..10305
repeat_region      /rpt_family="AluJo"
                    complement(10310..10348)
repeat_region      /rpt_family="MAR1"
                    10356..10643
repeat_region      /rpt_family="AluSx"
                    10841..10910

Query Match      81.7% Score 18.8; DB 9; Length 20838;
Best Local Similarity 90.9%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTTATTGATGGGATAGAGG 23
    |||||
Db 1078 ATGTTTGGATGGGATAGGG 1099

RESULT 9
AC139617
LOCUS      Homo sapiens 3 BAC RP11-722C17 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION  AC139617.6 GI:28557821
VERSION     HTG.
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 25901)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oeman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayale,M., Banke,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Huiy,K.S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mashiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E.,
Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Worley,K.C.
TITLE
JOURNAL
REFERENCE
AUTHORS
Worley,K.C.
TITLE
JOURNAL
REFERENCE
AUTHORS
Worley,K.C.
TITLE
JOURNAL
COMMENT
CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.
ANNOTATION OF FEATURES:
STSS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.
SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.
QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
ml.
Location/Qualifiers
1..25901
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-722C17"
repeat_region 14..65

```

Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 25901)
Worley,K.C.
Direct Submission
Submitted (07-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 25901)
Worley,K.C.
Direct Submission
Submitted (22-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 25901)
Worley,K.C.
Direct Submission
Submitted (25-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Feb 25, 2003 this sequence version replaced gi:28467052.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht

ml.
Location/Qualifiers

```

1..25901
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-722C17"
repeat_region 14..65

```



```

ACCESSION      AC133528
VERSION        AC133528.4
KEYWORDS       GI:27802070
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE      1 (bases 1 to 151700)
AUTHORS       Sulston,J.E. and Waterston,R.
TITLE         Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE       99063792
PUBMED        9847074

REFERENCE      2 (bases 1 to 151700)
AUTHORS       Ali,J., Haakenson,W., Dignan,G. and Doebber,A.
TITLE         The sequence of Homo sapiens BAC clone RP11-367H1
JOURNAL       Unpublished (2001)

REFERENCE      3 (bases 1 to 151700)
AUTHORS       Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (13-SEP-2002) Genome Sequencing Center, Washington
               University School of Medicine, 4444 Forest Park Parkway, St. Louis,
               MO 63108, USA

REFERENCE      4 (bases 1 to 151700)
AUTHORS       Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (23-OCT-2002) Genome Sequencing Center, Washington
               University School of Medicine, 4444 Forest Park Parkway, St. Louis,
               MO 63108, USA

REFERENCE      5 (bases 1 to 151700)
AUTHORS       Waterston,R.
TITLE         Direct Submission
JOURNAL       Submitted (27-NOV-2002) Department of Genetics, Washington
               University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE      6 (bases 1 to 151700)
AUTHORS       Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (18-JAN-2003) Genome Sequencing Center, Washington
               University School of Medicine, 4444 Forest Park Parkway, St. Louis,
               MO 63108, USA

REFERENCE      7 (bases 1 to 151700)
AUTHORS       Waterston,R.
TITLE         Direct Submission
JOURNAL       Submitted (19-FEB-2003) Department of Genetics, Washington
               University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT       On Jan 18, 2003 this sequence version replaced gi:25777548.
               ----- Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://genome.wustl.edu/gsc
               Contact: sapiens@wuston.wustl.edu
               ----- Summary Statistics
               -----
               Center project name: H_NH0367H01

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

This sequence is not the entire insert of the clone. This clone is overlapped by AC110299.

Discrepant bases between AC114730, AC132150 and clone sequence.

Data from AC132150 and AC114730 was used to finish this clone.

FEATURES

| Location/Qualifiers | Source |
|--------------------------|---------------|
| 1..151700 | |
| /organism="Homo sapiens" | |
| /mol_type="genomic DNA" | |
| /db_xref="taxon:9606" | |
| /chromosome="2" | |
| /map="2" | |
| /clone="RP11-367H1" | |
| /clone_lib="RPCI-11" | |
| 23..325 | |
| /rpt_family="Alu" | repeat_region |
| 896..1031 | repeat_region |
| /rpt_family="Alu" | |
| 1032..1160 | repeat_region |
| /rpt_family="Alu" | |
| 1161..1472 | repeat_region |
| /rpt_family="Alu" | |
| 1473..1655 | repeat_region |
| /rpt_family="Alu" | |
| 1656..1840 | repeat_region |
| /rpt_family="Alu" | |
| 1853..2023 | repeat_region |
| /rpt_family="Alu" | |
| 2038..2325 | repeat_region |
| /rpt_family="Alu" | |
| 2765..2905 | repeat_region |
| /rpt_family="L1" | |
| 3107..3437 | repeat_region |
| /rpt_family="L1" | |
| 3438..3740 | repeat_region |
| /rpt_family="Alu" | |
| 3749..4044 | repeat_region |
| /rpt_family="Alu" | |
| 4045..4275 | repeat_region |
| /rpt_family="L1" | |
| 4353..4653 | repeat_region |
| /rpt_family="Alu" | |
| 4657..4685 | repeat_region |
| /rpt_family="L1" | |
| 4686..4987 | repeat_region |
| /rpt_family="Alu" | |
| 4988..5288 | repeat_region |
| /rpt_family="L1" | |
| 5289..5596 | repeat_region |
| /rpt_family="Alu" | |
| 5597..6232 | repeat_region |
| /rpt_family="L1" | |
| 6239..6557 | repeat_region |
| /rpt_family="Alu" | |
| 6567..6864 | repeat_region |
| /rpt_family="Alu" | |
| 7357..7658 | repeat_region |
| /rpt_family="Alu" | |
| 7861..7983 | repeat_region |
| /rpt_family="L2" | |

```

repeat_region 7993. .8042
repeat_region /rpt_family="MER1_type"
repeat_region 8043. .8346
repeat_region /rpt_family="Alu"
repeat_region 8347. .8441
repeat_region /rpt_family="MER1_type"
repeat_region 8516. .8720
repeat_region /rpt_family="L2"
misc_feature 8903. .10912
repeat_region /notes="CpG island (tGC=73.0, o/e=0.80, #CpGs=221)"
repeat_region 9462. .9569
repeat_region /rpt_family="G-rich"
repeat_region 9573. .9748
repeat_region /rpt_family="G-rich"
repeat_region 9807. .9837
repeat_region /rpt_family="(CCCCG)n"
repeat_region 9906. .9945
repeat_region /rpt_family="GC_rich"
repeat_region 10026. .10057
repeat_region /rpt_family="GC_rich"
repeat_region 10682. .10987
repeat_region /rpt_family="GC_rich"
repeat_region 11024. .11321
repeat_region /rpt_family="Alu"
repeat_region 11324. .11487
repeat_region /rpt_family="Alu"
repeat_region 11488. .11793
repeat_region /rpt_family="Alu"
repeat_region 11794. .11962
repeat_region /rpt_family="Alu"
repeat_region 12409. .12521
repeat_region /rpt_family="Alu"
repeat_region 12620. .12914
repeat_region /rpt_family="Alu"
repeat_region 12918. .13211
repeat_region /rpt_family="Alu"
repeat_region 13329. .13404
repeat_region /rpt_family="L1"
repeat_region 13405. .13447
repeat_region /rpt_family="(CA)n"
repeat_region 13448. .13547
repeat_region /rpt_family="L1"
repeat_region 13548. .13781
repeat_region /rpt_family="Alu"
repeat_region 13782. .13843
repeat_region /rpt_family="L1"
repeat_region 13844. .14141
repeat_region /rpt_family="Alu"
repeat_region 14142. .14490
repeat_region /rpt_family="L1"
repeat_region 14491. .14790
repeat_region /rpt_family="Alu"

Query Match 81.7%; Score 18.8; DB 9; Length 151700;
Best Local Similarity 90.9%; Pred. No. 3.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ATGATTGATGGGATAGG 23
Db 42701 ATATTGATGGGAAAGG 42680

RESULT 12
AC102466
LOCUS AC102466 163947 bp DNA linear HTG 16-SEP-2004
DEFINITION Mus musculus chromosome 3 clone RP24-337A2 map 3, *** SEQUENCING IN
PROGRESS ***, 5 unordered pieces.
ACCESSION AC102466
VERSION AC102466.6 GI-52138860
KEYWORDS HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE

1 (bases 1 to 163947)
 Birren, B., Nusbaum, C., and Lander, E.
 Mus musculus chromosome 3, clone RP24-337A2
 unpublished
 2 (bases 1 to 163947)

1 (bases 1 to 163947)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
 Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 163947)

1 (bases 1 to 163947)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, M., Anderson, S., Atacchi, H.M., Barna, N., Bastien, V.,
 Bloom, T., Boguslavskiy, L., Boukhgalter, B., Canarata, J., Chang, J.,
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
 DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
 Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D.,
 Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
 Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,
 MacLean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
 McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
 Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R.,
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
 Zimmer, A. and Zody, M.

Direct Submission
 Submitted (16-SEP-2004) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 16, 2004 this sequence version replaced gi:51536822.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@broad.mit.edu
 ----- Project Information
 Center project name: L18861
 Center clone name: 337_A_2

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

REFERENCE

1 (bases 1 to 163947)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, M., Anderson, S., Atacchi, H.M., Barna, N., Bastien, V.,
 Bloom, T., Boguslavskiy, L., Boukhgalter, B., Canarata, J., Chang, J.,
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
 DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
 Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D.,
 Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
 Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,
 MacLean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
 McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
 Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R.,
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
 Zimmer, A. and Zody, M.

Direct Submission
 Submitted (16-SEP-2004) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 16, 2004 this sequence version replaced gi:51536822.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@broad.mit.edu
 ----- Project Information
 Center project name: L18861
 Center clone name: 337_A_2

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 42857: contig of 42857 bp in length
 * 42858 42957: gap of unknown length
 * 42958 55670: contig of 12713 bp in length
 * 55671 55770: gap of unknown length
 * 55771 70601: contig of 14831 bp in length
 * 70602 70701: gap of unknown length
 * 70702 96871: contig of 26170 bp in length
 * 96872 96971: gap of unknown length
 * 96972 163947: contig of 66976 bp in length.

FEATURES

source

Location/Qualifiers
 1. .163947
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="3"
 /map="3"
 /clone="RP24-337A2"
 /clone_lib="RPCI-24 Male Mouse BAC"

ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 163947;
 Best Local Similarity 90.9%; Pred. No. 3.4e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTGATGGGATAGAGG 23
 |||||

Db 42059 ATTATTGATGGGAGAGG 42080
 |||||

RESULT 13

AC151841

LOCUS

AC151841 167797 bp DNA linear HTG 06-OCT-2004
 Mus musculus chromosome 7 clone RP23-128P10, *** SEQUENCING IN
 PROGRESS ***, 63 unordered pieces.

ACCESSION

AC151841

VERSION

AC151841.1 GI:53828886

KEYWORDS

HTG; HTGS PHASE1.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Theria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Wilson, R.K.
 The sequence of Mus musculus clone
 1 (bases 1 to 167797)
 Unpublished
 2 (bases 1 to 167797)
 Wilson, R.K.
 Direct Submission
 Submitted (06-OCT-2004) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu
 ----- Project Information -----
 Center project name: M_BA0128P10
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Chemistry: Dye-terminator ET; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 132837 bases at least Q40
 Consensus quality: 141875 bases at least Q30
 Consensus quality: 146026 bases at least Q20

 * NOTE: This is a 'working draft' sequence. It currently

* consists of 63 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1310: contig of 1310 bp in length
 1410: gap of unknown length
 1411 2848: contig of 1438 bp in length
 2849 2948: gap of unknown length
 2949 4104: contig of 1156 bp in length
 4105 4204: gap of unknown length
 4205 5597: contig of 1393 bp in length
 5598 5697: gap of unknown length
 5698 7028: contig of 1231 bp in length
 7029 8365: contig of 1337 bp in length
 8366 8465: gap of unknown length
 8466 9878: contig of 1413 bp in length
 9879 11628: contig of 1650 bp in length
 11629 11728: gap of unknown length
 11729 13119: contig of 1391 bp in length
 13120 13219: gap of unknown length
 13220 14722: contig of 1503 bp in length
 14723 14822: gap of unknown length
 14823 16105: contig of 1283 bp in length
 16106 16205: gap of unknown length
 16206 17617: contig of 1412 bp in length
 17618 17717: gap of unknown length
 17719 19179: contig of 1462 bp in length
 19180 19279: gap of unknown length
 19280 20692: contig of 1413 bp in length
 20693 20792: gap of unknown length
 20793 22378: contig of 1586 bp in length
 22379 22478: gap of unknown length
 22479 23815: contig of 1337 bp in length
 23816 23915: gap of unknown length
 23916 25183: contig of 1268 bp in length
 25184 25283: gap of unknown length
 25284 26438: contig of 1155 bp in length
 26439 26538: gap of unknown length
 26539 28247: contig of 1709 bp in length
 28248 28347: gap of unknown length
 28348 29530: contig of 1183 bp in length
 29531 29630: gap of unknown length
 29631 32037: contig of 2407 bp in length
 32038 32137: gap of unknown length
 32138 33577: contig of 1440 bp in length
 33578 33677: gap of unknown length
 33678 35497: contig of 1820 bp in length
 35498 35597: gap of unknown length
 35598 37174: contig of 1577 bp in length
 37175 37274: gap of unknown length
 37275 38990: contig of 1716 bp in length
 38991 39090: gap of unknown length
 39091 40783: contig of 1693 bp in length
 40784 40883: gap of unknown length
 40884 42469: contig of 1586 bp in length
 42470 42569: gap of unknown length
 42570 44071: contig of 1502 bp in length
 44072 44171: gap of unknown length
 44172 45755: contig of 1584 bp in length
 45756 45855: gap of unknown length
 45856 48304: contig of 2449 bp in length
 48305 48404: gap of unknown length
 48405 49626: contig of 1222 bp in length
 49627 51219: contig of 1493 bp in length
 51220 51319: gap of unknown length
 51320 53580: contig of 2261 bp in length
 53581 53680: gap of unknown length

```

* 53681 55124: contig of 1444 bp in length
* 55125 55224: gap of unknown length
* 55226 56541: contig of 1317 bp in length
* 56542 56641: gap of unknown length
* 56642 58182: contig of 1541 bp in length
* 58183 58282: gap of unknown length
* 58283 61376: contig of 3094 bp in length
* 61377 61476: gap of unknown length
* 61477 63361: contig of 1885 bp in length
* 63362 63461: gap of unknown length
* 63462 65401: contig of 1940 bp in length
* 65402 65501: gap of unknown length
* 65502 67057: contig of 1556 bp in length
* 67058 67157: gap of unknown length
* 67158 69300: contig of 2143 bp in length
* 69301 69400: gap of unknown length
* 69401 71077: contig of 1677 bp in length
* 71078 71177: gap of unknown length
* 71178 73396: contig of 2219 bp in length
* 73397 73496: gap of unknown length
* 73497 74631: contig of 1135 bp in length
* 74632 74731: gap of unknown length
* 74732 76901: contig of 2170 bp in length
* 76902 77001: gap of unknown length
* 77002 79346: contig of 2345 bp in length
* 79347 79446: gap of unknown length
* 79447 81147: contig of 1701 bp in length
* 81148 81247: gap of unknown length
* 81248 85090: contig of 3843 bp in length
* 85091 85190: gap of unknown length
* 85191 88634: contig of 3444 bp in length
* 88635 88734: gap of unknown length
* 88735 91492: contig of 2758 bp in length
* 91493 91592: gap of unknown length
* 91593 95177: contig of 3585 bp in length
* 95178 95277: gap of unknown length
* 95279 100040: contig of 4763 bp in length
* 100041 100140: gap of unknown length
* 100141 105686: contig of 5546 bp in length
* 105687 105786: gap of unknown length
* 105787 111728: contig of 5942 bp in length
* 111729 111828: gap of unknown length
* 111829 116854: contig of 5026 bp in length
* 116855 116954: gap of unknown length
* 116955 123034: contig of 6080 bp in length
* 123035 123134: gap of unknown length
* 123135 128962: contig of 5828 bp in length
* 128963 129062: gap of unknown length
* 129063 133910: contig of 4848 bp in length
* 133911 134010: gap of unknown length
* 134011 139088: contig of 5078 bp in length
* 139089 139188: gap of unknown length
* 139189 145581: contig of 6393 bp in length
* 145582 145681: gap of unknown length
* 145682 153130: contig of 7449 bp in length
* 153131 153230: gap of unknown length
* 153231 159350: contig of 6120 bp in length
* 159351 159450: gap of unknown length
* 159451 167797: contig of 8347 bp in length.

```

FEATURES

```

source
1. .167797
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="7"
  /clone="RP23-128P10"
misc_feature
1. .1310
  /note="assembly_name:Contig14"
misc_feature
1411. .2848
  /note="assembly_name:Contig20"
misc_feature
2949. .4104
  /note="assembly_name:Contig22"
misc_feature
4205. .5597

```

```

misc_feature /note="assembly_name:Contig23"
5698. .6928
misc_feature /note="assembly_name:Contig25"
7029. .8365
misc_feature /note="assembly_name:Contig26"
8466. .9878
misc_feature /note="assembly_name:Contig27"
9979. .11628
misc_feature /note="assembly_name:Contig28"
11729. .13119
misc_feature /note="assembly_name:Contig31"
13220. .14722
misc_feature /note="assembly_name:Contig32"
14823. .16105
misc_feature /note="assembly_name:Contig34"
16206. .17617

```

Query Match 81.7%; Score 18.8; DB 2; Length 167797;

Best Local Similarity 90.9%; Pred. NO. 3.4e+02; Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTGTGGGATAGAGG 23

Db 43214 ATGTATTGTGGGATAGAGG 43235

RESULT 14

AC118744

LOCUS AC118744 171446 bp DNA linear HTG 24-FEB-2004
Mus musculus chromosome 6 clone RP24-182M11 map 6, *** SEQUENCING
IN PROGRESS ***, 7 unordered pieces.

ACCESSION AC118744

VERSION AC118744.5 GI:42761784

KEYWORDS HTG; HTGS; PHASE1; HTGS_FULLTOP; HTGS_ACTIVEPIN.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 171446)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus chromosome 6, clone RP24-182M11

Unpublished

2 (bases 1 to 171446)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faroo, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Govette, M., Graham, L., Grand-Pierre, N.,
Hagoe, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karas, A., Kells, C., LaRocque, K., Lamazares, R., C.,
Landers, T., Lechoczky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 171446)

REFERENCE

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K.,

Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-FEB-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 24, 2004 this sequence version replaced gi:28570380.
All repeats were identified using RepeatMasker:
Smith, A.F.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L25190

Center Clone name: 182_M_11

* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 33708: contig of 33708 bp in length
* 33709 33808: gap of 100 bp
* 33809 43123: contig of 9315 bp in length
* 43124 43223: gap of 100 bp
* 43224 46026: contig of 2803 bp in length
* 46027 46126: gap of 100 bp
* 46127 81380: contig of 35254 bp in length
* 81381 81480: gap of 100 bp
* 81481 139660: contig of 58180 bp in length
* 139661 139760: gap of 100 bp
* 139761 146196: contig of 6436 bp in length
* 146197 146296: gap of 100 bp
* 146297 171446: contig of 25150 bp in length.

FEATURES

Source

1. 171446
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosomes="6"
/map="6"
/clone="RP24-182M11"
/clone_lib="RP24-182M11 Male Mouse BAC"

ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 171446;
Best Local Similarity 90.9%; Pred. No. 3.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CATGTAATTCATCGGGATAGAG 22

|||||
Db 136493 CATGTAATTCGCTAGGGATAGAG 136514

RESULT 15
AC113270/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC113270 175280 bp DNA linear ROD 24-OCT-2003
Mus musculus chromosome 7, clone RP23-375O10, complete sequence.
AC113270
HTG
Mus musculus (house mouse)
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 175280)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus chromosome 7, clone RP23-375O10
Unpublished
2 (bases 1 to 175280)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Canarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 175280)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (28-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 175280)
Birren, B., Nusbaum, C. and Lander, E.
Direct Submission
Submitted (10-SEP-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (bases 1 to 175280)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,

Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., Matthews, C., McCarthy, M., Macdonald, P., Major, J., Manning, J., Menga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
6 (bases 1 to 175280)

TITLE

JOURNAL

REFERENCE

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., Matthews, C., McCarthy, M., Macdonald, P., Major, J., Manning, J., Menga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-OCT-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2003 this sequence version replaced gi:33147222.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L22700

Center clone name: 375_O_10

FEATURES

source

----- Location/Qualifiers

1..175280

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="7"

/map="7"

/clone="RP23-375O10"

/clone_lib="RPCI-23 Female Mouse BAC"

1..9643

/note="wgs end extension"

clone_end:SP6"

116..122

/note="<30 qual SNGL region"

complement(323..3159)

misc_feature

unsure

repeat_region

repeat_region
/rpt_family="L1_RN"
3167..3204
/rpt_family="AT_rich"
3258..3278
/rpt_family="AT_rich"
complement(3404..3670)
/rpt_family="Lx8"
3724..3758
/rpt_family="(TAGA)n"
complement(3871..4067)
/rpt_family="UBR1A"
complement(4099..4801)
/rpt_family="Lx7"
4800..5537
/rpt_family="Lx9"
5823..5869
/rpt_family="(TC)n"
5870..5932
/rpt_family="(GA)n"
complement(7112..7646)
/rpt_family="RLTR11A"
7688..7910
/rpt_family="B4A"
7928..8000
/rpt_family="(TC)n"
8100..8188
/rpt_family="Lx9"
complement(8230..9007)
/rpt_family="Lx2"
9644..9649
/note="clone boundary
clone_end:SP6
site:EcoRI"
complement(9674..9934)
/rpt_family="LIR2_RN"
9935..9965
/rpt_family="(TTG)n"
complement(9966..10554)
/rpt_family="LIR2_RN"
10554..10613
/rpt_family="Lx9"
10752..10814
/rpt_family="(TG)n"
10923..10950
/rpt_family="AT_rich"
complement(11697..11792)
/rpt_family="MIR"
12030..12057
/rpt_family="(TTAA)n"
12232..12275
/rpt_family="(TTCTCC)n"
12504..12637
/rpt_family="B1P"
12702..12850
/rpt_family="B1_MM"
12868..12942

Query Match

Best Local Similarity 81.7%; Score 18.8; DB 10; Length 175280;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 CATGTATTTGATGGGATAGAG 22

|||||

DB 82140 CATGCATTTGATGGGATAGAG 82119

|||||

Search completed: August 13, 2005, 05:04:08

Job time : 740.234 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 12, 2005, 22:21:57 ; Search time 184.245 Seconds
(without alignments)
738.985 Million cell updates/sec

Title: US-10-673-854-1

Perfect score: 23

Sequence: 1 catgtatttgatgggagatagag 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 18.8 | 81.7 | 5651 | 4 ABA08791 | Aba08791 Human dJ8 |
| c 2 | 18.8 | 81.7 | 5653 | 4 AAH99780 | Aah99780 Human pro |
| c 3 | 18.8 | 81.7 | 5653 | 4 AAi60854 | Aai60854 Human pol |
| 4 | 18.8 | 81.7 | 5654 | 4 AAH99496 | Aah99496 Human pro |
| 5 | 18.8 | 81.7 | 5665 | 4 ABA08796 | Aba08796 Human dJ8 |
| 6 | 18.4 | 80.0 | 493 | 10 ADE81837 | Ade81837 Arabidops |
| 7 | 18.4 | 80.0 | 1292 | 3 AAC33276 | Aac33276 Arabidops |
| 8 | 18.4 | 80.0 | 3050 | 13 ADRO6686 | Adro6686 Full leng |
| 9 | 18.2 | 79.1 | 1630 | 8 ACC00653 | Acc00653 Argemone |
| 10 | 18.2 | 79.1 | 1630 | 10 ADC23534 | Adc23534 cDNA enco |
| c 11 | 18.2 | 79.1 | 1917 | 6 ABV75719 | Abv75719 Human B-C |
| c 12 | 18.2 | 79.1 | 22477 | 11 ACN44910 | Acn44910 Human gen |
| c 13 | 18.2 | 79.1 | 29111 | 11 ACN44424 | Acn44424 Mouse gen |
| c 14 | 18 | 78.3 | 2286 | 13 ADS46987 | Ads46987 Bacterial |
| c 15 | 17.8 | 77.4 | 238 | 5 ABV12937 | Abv12937 Human pro |
| c 16 | 17.8 | 77.4 | 238 | 5 ABV03768 | Abv03768 Human pro |
| c 17 | 17.8 | 77.4 | 416 | 5 ABV34062 | Abv34062 Human pro |
| c 18 | 17.8 | 77.4 | 745 | 4 AAi93090 | Aai93090 Human pol |
| c 19 | 17.8 | 77.4 | 1054 | 6 ABQ30767 | Abq30767 Oligonuc |
| 20 | 17.8 | 77.4 | 1054 | 6 ABQ30766 | Abq30766 Oligonuc |

| | | | | | |
|------|------|------|--------|---------------|--------------------|
| c 21 | 17.8 | 77.4 | 2000 | 6 ABZ16271 | Abz16271 Arabidops |
| 22 | 17.8 | 77.4 | 5040 | 6 ABL34404 | Ab134404 Human imm |
| 23 | 17.8 | 77.4 | 5975 | 6 ABL32236 | Ab132236 Human imm |
| 24 | 17.8 | 77.4 | 11046 | 6 ABK31536 | Abk31536 Signal tr |
| 25 | 17.2 | 74.8 | 33 | 6 ABA95678 | Aba95678 Human int |
| c 26 | 17.2 | 74.8 | 648 | 10 ADD33537 | Add33537 Mouse mit |
| c 27 | 17.2 | 74.8 | 957 | 6 ABN98549 | Abn98549 Arabidops |
| 28 | 17.2 | 74.8 | 1026 | 4 AAH52942 | Aah52942 S. epider |
| 29 | 17.2 | 74.8 | 2250 | 4 AAH52372 | Aah52372 S. epider |
| 30 | 17.2 | 74.8 | 2430 | 8 ACA46750 | Ac46750 Prokaryot |
| 31 | 17.2 | 74.8 | 2448 | 6 ABN91314 | Abn91314 Staphyloc |
| c 32 | 17.2 | 74.8 | 2448 | 13 ADS01065 | Ads01065 Staphyloc |
| c 33 | 17.2 | 74.8 | 3427 | 4 AAH54925 | Aah54925 S. epider |
| 34 | 17.2 | 74.8 | 4198 | 4 AAH54240 | Aah54240 S. epider |
| 35 | 17.2 | 74.8 | 8700 | 6 ABL33003 | Ab133003 Human imm |
| c 36 | 17.2 | 74.8 | 11622 | 6 ABL32674 | Ab132674 Human imm |
| c 37 | 17.2 | 74.8 | 47745 | 11 ACN44160 | Acn44160 Mouse gen |
| 38 | 17.2 | 74.8 | 110000 | 12 ADQ97331.2 | Continuation (3 of |
| 39 | 17.2 | 74.8 | 205388 | 12 ADQ97560 | Adq97560 Mouse can |
| 40 | 16.8 | 73.0 | 367 | 8 ABX49588 | Abx49588 Bovine ES |
| c 41 | 16.8 | 73.0 | 387 | 4 AAi85735 | Aai85735 Human pol |
| c 42 | 16.8 | 73.0 | 520 | 6 AAK99071 | Aak99071 520nt DNA |
| c 43 | 16.8 | 73.0 | 520 | 6 AAK99114 | Aak99114 DNA encod |
| c 44 | 16.8 | 73.0 | 583 | 6 ABK53412 | Abk53412 Human eos |
| 45 | 16.8 | 73.0 | 5757 | 4 ABL08602 | Ab108602 Drosophil |

ALIGNMENTS

RESULT 1

ABA08791

ID ABA08791 standard; cDNA; 5651 BP.

AC ABA08791;

DT 11-JAN-2002 (first entry)

DE Human dJ889M15.3 homologue-encoding cDNA, SEQ ID NO:567.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytotatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antitumor; antitumor; ss.

XX Homo sapiens.

XX WO200157188-A2.

PD 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US03800.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00560875.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457740/49.

XX P-PSDB; ABL11547.

PT Human proteins and DNA encoding sequences useful for preventing, treating

PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
XX and cancer.

PS Claim 1; Page 582-584; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention

XX Sequence 5651 BP; 1077 A; 1633 C; 1680 G; 1261 T; 0 U; 0 Other;

Query Match 81.7%; Score 18.8; DB 4; Length 5651;
Best Local Similarity 90.9%; Pred. No. 1e-02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTTCATGGGATAGAGG 23

DB 1556 ATATATTTCATGGGAAAGAGG 1577

RESULT 2

AAH99780/c

ID AAH99780 standard; cDNA; 5653 BP.

XX AC AAH99780;

XX 16-OCT-2001 (first entry)

XX Human protein encoding cDNA sequence SEQ ID NO:615.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;

XX antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX antibacterial; endocrine; cardiant; central nervous system; virucide;
XX anti-HIV; fungicide; antitumor; cardiovascular; antianaemic; anaemia;
XX antiaggregant; haemostatic; vulnerary; antileuc; osteopathic; eczema;
XX dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
XX neuroprotective; nootropic; antiparkinsonian; infection;
XX immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
XX antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;

KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ps.

XX Homo sapiens.

PN WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US035017.

XX 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457603/49.

XX P-PSDB; AAM25839.

XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX Claim 1; Page 642-644; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and cells
CC they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antitumor;
CC cardiovascular; antianaemic; antileuc; haemostatic; vulnerary;
CC antileuc; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, ulcers, eczema,
CC osteoporosis, severe combined immunodeficiency, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders

XX Sequence 5653 BP; 1262 A; 1680 C; 1632 G; 1079 T; 0 U; 0 Other;

Query Match 81.7%; Score 18.8; DB 4; Length 5653;
Best Local Similarity 90.9%; Pred. No. 1e-02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTTCATGGGATAGAGG 23

DB 4098 ATATATTTCATGGGAAAGAGG 4077

RESULT 3

AAI60854/c

ID AAI60854 standard; cDNA; 5653 BP.

XX AC AAI60854;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4843.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX OS Homo sapiens.
 XX KW WO200153312-A1.
 XX PD 26-JUL-2001.
 XX PF 26-DEC-2000; 2000WO-US034263.
 XX PR 23-DEC-1999; 99US-00471275.
 XX PR 21-JAN-2000; 2000US-0048725.
 XX PR 25-APR-2000; 2000US-00552317.
 XX PR 20-JUN-2000; 2000US-00598042.
 XX PR 19-JUL-2000; 2000US-00620312.
 XX PR 03-AUG-2000; 2000US-00653450.
 XX PR 14-SEP-2000; 2000US-00662191.
 XX PR 19-OCT-2000; 2000US-00693036.
 XX PR 29-NOV-2000; 2000US-00727344.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QH;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX DR WPI; 2001-442253/47.
 XX DR P-PSDB; AAM41698.
 XX PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX PS Claim 1; SEQ ID NO 4843; 10078pp; English.
 XX CC The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX SQ Sequence 5653 BP; 1262 A; 1680 C; 1632 G; 1079 T; 0 U; 0 Other;
 Query Match 81.7%; Score 18.8; DB 4; Length 5653;
 Best Local Similarity 90.9%; Pred. No. 1e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 ATGTATTGATGGGATAGG 23
 Db 4098 ATATATTGATGGGAAGG 4077
 RESULT 4
 AAH99496
 ID AAH99496 standard; cDNA; 5654 BP.
 XX AC AAH99496;
 XX

DT 16-OCT-2001 (first entry)
 XX Human protein encoding cDNA sequence SEQ ID NO:331.
 DE Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 XX antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 XX antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; anaemia;
 KW antiaggregant; haemostatic; vulnary; antilucer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder; ss.
 OS Homo sapiens.
 XX WO200153455-A2.
 XX PD 26-JUL-2001.
 XX PF 22-DEC-2000; 2000WO-US035017.
 XX PR 23-DEC-1999; 99US-00471275.
 XX PR 21-JAN-2000; 2000US-00488725.
 XX PR 25-APR-2000; 2000US-00552317.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Drmanac RT;
 XX DR WPI; 2001-457603/49.
 XX DR P-PSDB; AAM25555.
 XX PT Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
 XX PS Claim 1; Page 444-446; 1217pp; English.
 XX CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and cells
 CC they are expressed in, such as: antinflammatory; antirheumatic;
 CC antarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; anaemia; antiaesthetic; haemostatic; vulnary;
 CC antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders
 XX SQ Sequence 5654 BP; 1078 A; 1631 C; 1682 G; 1263 T; 0 U; 0 Other;
 Query Match 81.7%; Score 18.8; DB 4; Length 5654;
 Best Local Similarity 90.9%; Pred. No. 1e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 ATGTATTGATGGGATAGG 23

Db 1559 ATATATTGATGGGGAAGG 1580
|||||
RESULT 5
ABA08796
ID ABA08796 standard; cDNA; 5665 BP.
AC ABA08796;
XX
XX
11-JAN-2002 (first entry)
XX
XX Human dJ889M15.3 homologue-encoding cDNA, SEQ ID NO:572.
XX
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer; ss.
XX
XX Homo sapiens.
XX
XX WO200157188-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US003800.
XX
XX 03-FEB-2000; 2000US-00495914.
PR 27-APR-2000; 2000US-00560875.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX WPI; 2001-457740/49.
DR P-PSDB; ABB11552.
XX
XX Human proteins and DNA encoding sequences useful for preventing, treating
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
PT and cancer.
XX
XX Claim 1; Page 586-587; 1963pp; English.
XX
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell

CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness.
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
XX novel human polypeptide of the invention
SQ Sequence 5665 BP; 1077 A; 1634 C; 1687 G; 1267 T; 0 U; 0 Other;
Query Match 81.7%; Score 18.8; DB 4; Length 5665;
Best Local Similarity 90.9%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 ATGTATTGATGGGATAGAGG 23
Db 1567 ATATATTGATGGGGAAGG 1588
|||||
RESULT 6
ADE81837
ID ADE81837 standard; cDNA; 493 BP.
XX
XX ADE81837;
XX
XX 29-JAN-2004 (first entry)
XX
XX Arabidopsis thaliana expressed polynucleotide seq id 608.
DE
XX
KW genetically modified organism; transgenic organism; plant;
KW inhibitor testing; activator testing; modifier testing; fungicide;
KW insecticide; genetic function; genetic regulation; cellular metabolism;
KW gene; ss.
XX
XX Arabidopsis thaliana.
XX
XX US2003115639-A1.
XX
XX 19-JUN-2003.
XX
XX 26-JAN-2001; 2001US-00770961.
XX
XX 27-JAN-2000; 2000US-0178466P.
XX
XX (GORL/) GORLACH J.
PA (ANYV/) AN Y.
PA (HAML/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUYV/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A. V.
PA (MATH/) MATHAW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX

Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
Garcia CA, Krickler M, Slater T, Davis KR, Allen K, Hoffman N;
Hurban P;
XX
WPI; 2003-810930/76.
XX
Novel Arabidopsis thaliana nucleic acids useful for generating
genetically modified transgenic organisms, for screening biologically
active agents such as fungicides, insecticides.
XX
Claim 1; SEQ ID NO 608; 44pp; English.
XX
The invention describes a nucleic acid (I) comprising a sequence capable
of hybridising under stringent conditions to any one of 999 fully defined
Arabidopsis thaliana sequences (I) as given in specification e.g., 360,
1137, 455, 219, 472 nucleotides, etc, or its fragment. (I) is useful as a
hybridisation probe to complementary molecules in a cDNA library. (I) is
also useful for generating genetically modified and transgenic organisms
usually plant cells and plants. A protein encoded by (I) is useful in
screening assays to determine the effect of candidate inhibitors,
activators or modifiers of the gene product. The protein is also useful
for screening biologically active agents e.g., fungicides and
insecticides. A genetically modified cell, comprising an exogenous
nucleic acid, where the nucleic acid comprises transcription regulatory
sequences operably linked to a sequence capable of hybridising under
stringent conditions to (I) is useful in the study of genetic function
and regulation, for alteration of the cellular metabolism and for
screening compounds that may affect the biological function of the gene
or gene product. This sequence represents an Arabidopsis thaliana
polynucleotide of the invention.

Sequence 493 BP; 158 A; 88 C; 121 G; 126 T; 0 U; 0 Other;

Query Match 80.0%; Score 18.4; DB 10; Length 493;
Best Local Similarity 95.0%; Pred No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY 4 GTATTTTCATGGGGATAGAGG 23
| | | | | | | | | | | | | |
Db 162 GCATTTTCATGGGGATAGAGG 181

RESULT 7
AAC33276
ID AAC33276 standard; DNA; 1292 BP.
XX AAC33276;
AC
XX
XX 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 2433.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway; metabolic pathway;
promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
FN EF1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.

```
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144332P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.

PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159637P.
PR 16-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161408P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      80.0%; Score 18.4; DB 3; Length 1292;
Best Local Similarity 95.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GTATTGATGGGATAGAGG 23
   |||||
Db 979 GCATTGATGGGATAGAGG 998

RESULT 8
ADR06686
ID ADR06686 standard; cDNA; 3050 BP.
XX
AC ADR06686;
XX
DT 04-NOV-2004 (first entry)
XX
DE Full length human cDNA useful for treating neurological disease Seq 192.
KW gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KW osteoporosis; neurological disease; Alzheimer's disease;
KW Parkinson's disease; dementia; short memory; cancer;
KW sense or motor function; emotional reaction; fear response; panic;
KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW tranqulliser.
XX
OS Homo sapiens.
XX
PN EP1447413-A2.
XX
PD 18-AUG-2004.
XX
PP 12-FEB-2004; 2004EP-00003145.
XX
PR 14-FEB-2003; 2003JP-00102207.
PR 09-MAY-2003; 2003JP-00131452.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
XX
DR WPI; 2004-583265/P.
XX
P-PSDB; ADR08642.
XX
PT New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
```

PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.

XX Claim 1; SEQ ID NO 192; 2686pp; English.

PS This invention relates to novel, isolated full length human cDNA
 CC molecules and the encoded proteins thereof. Specifically, it refers to
 CC cDNA clones obtained by an oligo-capping method, where none of these
 CC clones are identical to any known human mRNAs. The present invention
 CC describes an immunosay to identify agonists and antagonists, as well as
 CC antibodies, antisense molecules and siRNAs that can all be used to bind
 CC to and modulate expression of the cDNA molecules. As such, these
 CC molecules are useful for diagnostic markers or therapeutic targets for
 CC the various diseases or morbid states. In particular, they are useful in
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
 CC disease, Parkinson's disease, dementia, short memory and various cancers,
 CC as well as for maintaining equilibrium of sense or motor function, and
 CC for treating emotional reaction, fear response and panic. Accordingly,
 CC they exhibit osteoprotective, neuroprotective, nootropic, antiparkinsonian,
 CC cyostatic and tranquiliser activities. This polynucleotide is a full
 CC length human cDNA sequence of the invention. NOTE: This sequence is not
 CC given in the sequence listing of the specification but can be obtained on
 CC CD-ROM from the European Patent Office, Vienna Sub-office.

XX SQ Sequence 3050 BP; 978 A; 562 C; 645 G; 865 T; 0 U; 0 Other;
 Query Match 80.0%; Score 18.4; DB 13; Length 3050;
 Best Local Similarity 95.0%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATGTATTGATGGGATAGA 21

Db 1616 ATGTATTGATGGGATACA 1635

RESULT 9

ACC00653

ID ACC00653 standard; cDNA; 1630 BP.

XX AC ACC00653;

XX DT 16-MAY-2003 (first entry)

XX DE Argemone mexicana oil trait related cDNA sequence SEQ ID NO:55.

XX KW Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;

XX KW receptor-like protein kinase; mitogen activated protein kinase; oil;

XX KW LIPIS-like transcription factor caleosin; ATP citrate lyase; SNF1;

XX KW CKC-like transcription factor; antisense inhibition; co-suppression;

XX KW transgenic plant; gene; ss.

XX OS Argemone mexicana.

XX XX

XX PN WO2003002751-A2.

XX PD 09-JAN-2003.

XX PF 27-JUN-2002; 2002WO-US020152.

XX PR 29-JUN-2001; 2001US-0301913P.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX PA (PION-) PIONEER HI-BRED INT INC.

XX PI Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvell LT;

XX PI Jones TJ, Kinney AJ, Klein TM, Li C, Oliveira IC, Sakai H, Shen B;

XX PI Tarczynski MC;

XX DR WPI; 2003-201509/19.

XX DR P-PSDB; ABR40618.

XX XX

XX PT Novel nucleotide fragment encoding polypeptides having receptor-like

XX PT protein kinase activity, caleosin-like activity, useful for altering oil

XX PT phenotypes in plants such as sunflower, coconut, soybean, wheat and rice.

XX

PS

CC Claim 18; Page 129; 542pp; English.

CC The present invention describes an isolated nucleotide fragment (I)

CC comprising a nucleic acid sequence (NS) chosen from a NS encoding a

CC polypeptide (PP) having receptor-like protein kinase activity, mitogen

CC activated protein (MAP)-kinase activity, LIPIS-like transcription factor

CC activity, caleosin-like activity, ATP citrate lyase activity, SNF1-like

CC activity and CKC-like transcription factor activity. Also described: (1)

CC complement (II) of (I); (2) a chimeric construct (III) comprising (I) or

CC (II), operably linked to a regulatory sequence; (3) a plant (IV)

CC comprising (III) in its genome; (4) seeds (V) obtained from (IV); and (5)

CC oil obtained from (V). (I) or its part can be used in antisense

CC inhibition or co-suppression in a transformed plant. (III) is useful for

CC altering the oil phenotype in a plant such as corn, soybean, wheat, rice,

CC canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for

CC creating transgenic plants having altered lipid profiles. (I) can also be

CC used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to

CC ABR40879 represent sequences used in the exemplification of the present

CC invention

XX SQ

Sequence 1630 BP; 497 A; 318 C; 375 G; 440 T; 0 U; 0 Other;

Query Match 79.1%; Score 18.2; DB 8; Length 1630;

Best Local Similarity 87.0%; Pred. No. 1.7e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CATGTATTGATGGGATAGAGG 23

Db 1328 CTGTGTTTGTGCGGAAGAGG 1350

RESULT 10

ADC23534

ID ADC23534 standard; cDNA; 1630 BP.

XX AC ADC23534;

XX DT 18-DEC-2003 (first entry)

XX DE cDNA encodes protein involved in altering plant oil phenotype (SeqID 39).

XX KW ss; gene; oil phenotype; plant; breeding; Hap transcription factor;

XX KW lipid biosynthesis; Hap2-like; Haps-like; Hap3/Lec1-like; prickly poppy.

XX OS Argemone mexicana.

XX XX

XX PN WO2003001902-A2.

XX PD 09-JAN-2003.

XX PF 27-JUN-2002; 2002WO-US022086.

XX PR 29-JUN-2001; 2001US-0301913P.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX PA (PION-) PIONEER HI-BRED INT INC.

XX PI Allen WB, Cahoon RE, Famodu OO, Harvell LT, Helentjaris TG, Li C;

XX PI Lowe KS, Oliveira IC, Shen B, Tarczynski MC;

XX DR WPI; 2003-210187/20.

XX DR P-PSDB; ADC23535.

XX XX

XX PT Novel nucleotide fragment encoding polypeptides having Hap2, Hap5 or

XX PT Hap3/Lec1-like activity useful for altering oil phenotypes in plants such

XX PT as sunflower, coconut, soybean, wheat and rice.

XX PS

Claim 19; SEQ ID NO 39; 202pp; English.

XX The present invention relates to a novel method for altering oil

XX phenotype in plants through the controlled expression of selective genes,

XX useful in plant breeding. Specifically, it refers to the isolated nucleic

CC acid sequences of cDNA clones that encode polypeptides homologous to
CC proteins of the Hap transcription factor family. Hap proteins have been
CC shown to be regulators of several important genes involved in lipid
CC biosynthesis, such as fatty acid synthase, and are also involved in
CC diverse processes from cell cycle regulation to metabolic control and
CC homeostasis. The present invention describes proteins having Hap2-like or
CC Hap5-like transcription factor activity, or alternatively Hap3/Lec1-like
CC activity, which can be used to transform plants such that it is possible
CC to select those with an altered phenotype. This polynucleotide sequence
CC is derived from a cDNA clone that encodes a protein homologous to those
CC involved in altering plant oil phenotypes, in an exemplification of the
CC invention.

XX SQ Sequence 1630 BP; 497 A; 318 C; 375 G; 440 T; 0 U; 0 Other;

Query Match 79.1%; Score 18.2; DB 10; Length 1630;
Best Local Similarity 87.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CATGTATTTCATGGGGATAGG 23
Db 1328 CTTGTGTTTCATGGGGAAAGG 1350

RESULT 11

ABV75719/c

ID ABV75719 standard; cDNA; 1917 BP.

XX AC

XX AC ABV75719;

XX DT

XX 29-JAN-2003 (first entry)

XX DE

XX Human B-class complementary phosphatidylol glycan 11.33 cDNA.

XX KW

XX Human; B-class complementary phosphatidylol glycan 11.33; infection;

XX KW

XX diabetes; tumour; gene; ss.

XX OS

XX Homo sapiens.

XX FH

XX Key Location/Qualifiers

XX CDS 1064..1375

XX FT /*tag= a

XX FT /product= "B-class complementary phosphatidylol glycan

XX FT 11.33"

XX PN

XX CN1351038-A.

XX PD

XX 29-MAY-2002.

XX PF

XX 26-OCT-2000; 2000CN-00125794.

XX PR

XX 26-OCT-2000; 2000CN-00125794.

XX PA

XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX PI

XX Mao Y, Xie Y;

XX DR

XX WPI; 2002-619861/67.

XX DR

XX P-PSDB; ABP57775.

XX XX

XX Claim 6; Page 25-26 (Disclosure); 33pp; Chinese.

XX PS

XX The invention relates to the novel human B-class complementary

XX phosphatidylol glycan 11.33, and the polynucleotide encoding it. The

XX polypeptide is useful for treating diseases such as infections, diabetes,

XX and tumour. The present sequence encodes the human B-class complementary

XX Phosphatidylol glycan 11.33 of the invention

XX XX

XX Sequence 1917 BP; 561 A; 371 C; 418 G; 567 T; 0 U; 0 Other;

Query Match 79.1%; Score 18.2; DB 6; Length 1917;
Best Local Similarity 87.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CATGTATTTCATGGGGATAGG 23
Db 83 CTTGTATCTGATGGGGAGAGG 61

RESULT 12

ACN44910

ID ACN44910 standard; DNA; 22477 BP.

XX AC

XX ACN44910;

XX XX

XX 18-NOV-2004 (first entry)

XX DT

XX Human genomic sequence HCG22355.

XX DE

XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.

XX KW

XX Homo sapiens.

XX OS

XX WO2003073826-A2.

XX PN

XX 12-SEP-2003.

XX PD

XX 28-FEB-2003; 2003WO-US006235.

XX PF

XX 01-MAR-2002; 2002US-00087192.

XX PR

XX (SAGR-) SAGRES DISCOVERY.

XX PA

XX Morris DW;

XX PI

XX WPI; 2003-328604/31.

XX DR

XX Claim 1; SEQ ID NO 1594; Opp; English.

XX PS

XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published

XX SQ Sequence 22477 BP; 5890 A; 5195 C; 5519 G; 5873 T; 0 U; 0 Other;

Query Match 79.1%; Score 18.2; DB 11; Length 22477;
Best Local Similarity 87.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CATGTATTTCATGGGGATAGG 23
Db 16428 CATATAATTGAGGGGATAGG 16450

RESULT 13

ACN44424/c

ID ACN44424 standard; DNA; 29111 BP.

XX XX

AC ACN44424;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Mouse genomic sequence MCG19046.
 XX
 KW Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
 XX
 OS Mus musculus.
 XX
 XX WC2003073826-A2.
 PN
 XX
 PD 12-SEP-2003.
 XX
 XX 28-FEB-2003; 2003WO-US006235.
 PF
 XX
 PR 01-MAR-2002; 2002US-00087192.
 XX
 XX (SAGR-) SAGRES DISCOVERY.
 PA
 XX
 PI Morris DW;
 XX
 XX WPI; 2003-328604/31.
 DR
 XX
 PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 PT comprises a nucleotide sequence.
 XX
 PS Claim 1; SEQ ID NO 865; Opp; English.
 XX
 CC The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CA coding
 CC sequence. Note: This patent is an equivalent to basic patent
 CC US2002182586A1, for which no sequence data was published
 XX
 SQ Sequence 29111 BP; 6425 A; 7887 C; 7445 G; 6494 T; 0 U; 860 Other;
 Query Match 79.1%; Score 18.2; DB 11; Length 29111;
 Best Local Similarity 79.0%; Pred. No. 2.5e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 CATGTATTTCATGGGATAGAGG 23
 DB 22398 CATGTATGGATGGGGAGAGG 22376
 RESULT 14
 ID ADS46987/c
 XX ADS46987 standard; cDNA; 2286 BP.
 XX
 AC ADS46987;
 XX
 XX 02-DEC-2004 (first entry)
 DT
 XX Bacterial polynucleotide #1730.
 DE
 XX Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.
 XX

OS Bacteria.
 XX
 PN US2003233675-A1.
 XX
 PD 18-DEC-2003.
 XX
 XX 20-FEB-2003; 2003US-00369493.
 PF
 XX 21-FEB-2002; 2002US-0360039P.
 PR
 XX (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX WPI; 2004-061375/06.
 DR
 XX
 PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 PT
 XX
 PS Claim 1; SEQ ID NO 25417; 122pp; English.
 XX
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 2286 BP; 697 A; 474 C; 512 G; 603 T; 0 U; 0 Other;
 Query Match 78.3%; Score 18; DB 13; Length 2286;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ATGTATTTCATGGGATA 19
 DB 656 ATGTATTTCATGGGATA 639
 RESULT 15
 ID ABV12937/c
 XX ABV12937 standard; cDNA; 238 BP.
 XX
 AC ABV12937;
 XX
 XX 13-SEP-2002 (first entry)
 DT
 XX Human prostate expression marker cDNA 12928.
 DE
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynanamic marker;
 KW

KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
XX 23-AUG-2001.
XX
PD
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX
XX Schlegel R, Endege WO, Monahan JE;
PI
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 2134; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 238 BP; 79 A; 60 C; 36 G; 63 T; 0 U; 0 Other;

Query Match 77.4%; Score 17.8; DB 5; Length 238;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TGTATTGTGGGGATAGGG 23
Db 54 TGTATTGTGGGGATAGGG 34

Search completed: August 13, 2005, 04:14:28
Job time : 190.245 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 03:33:32 ; Search time 56.5213 Seconds
(without alignments)
665.844 Million cell updates/sec

Title: US-10-673-854-1
Perfect score: 23
Sequence: 1 catgtatttgatggggatagagg 23

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 18.8 | 81.7 | 93364 | 4 | US-09-949-016-14890 |
| 2 | 17.8 | 77.4 | 89689 | 4 | US-09-949-016-13089 |
| 3 | 17.8 | 77.4 | 157644 | 4 | US-09-949-016-16179 |
| 4 | 17.8 | 77.4 | 157644 | 4 | US-09-949-016-16179 |
| 5 | 17.2 | 74.8 | 1026 | 4 | US-09-949-016-16180 |
| 6 | 17.2 | 74.8 | 2250 | 4 | US-09-710-279-1277 |
| 7 | 17.2 | 74.8 | 2448 | 3 | US-09-134-001C-777 |
| 8 | 17.2 | 74.8 | 3427 | 4 | US-09-710-279-4289 |
| 9 | 17.2 | 74.8 | 4198 | 4 | US-09-710-279-3604 |
| 10 | 16.8 | 73.0 | 71251 | 4 | US-09-949-016-15332 |
| 11 | 16.6 | 72.2 | 209 | 4 | US-09-313-294A-765 |
| 12 | 16.6 | 72.2 | 211 | 4 | US-09-513-999C-22576 |
| 13 | 16.6 | 72.2 | 601 | 4 | US-09-949-016-64747 |
| 14 | 16.6 | 72.2 | 627 | 3 | US-09-328-111-93 |
| 15 | 16.6 | 72.2 | 3880 | 4 | US-09-221-017B-1071 |
| 16 | 16.6 | 72.2 | 25274 | 4 | US-09-949-016-16682 |
| 17 | 16.6 | 72.2 | 25751 | 4 | US-09-949-016-13631 |
| 18 | 16.4 | 71.3 | 1917 | 4 | US-09-134-000C-1459 |
| 19 | 16.4 | 71.3 | 103712 | 4 | US-09-949-016-13058 |
| 20 | 16.4 | 71.3 | 236341 | 4 | US-09-949-016-13978 |
| 21 | 16.2 | 70.4 | 277 | 4 | US-09-313-294A-130 |
| 22 | 16.2 | 70.4 | 307 | 4 | US-09-313-294A-6619 |
| 23 | 16.2 | 70.4 | 441 | 4 | US-09-107-532A-2725 |
| 24 | 16.2 | 70.4 | 601 | 4 | US-09-949-016-205970 |
| 25 | 16.2 | 70.4 | 912 | 4 | US-09-107-532A-833 |
| 26 | 16.2 | 70.4 | 1941 | 4 | US-09-543-681A-2668 |
| 27 | 16.2 | 70.4 | 2250 | 4 | US-09-902-540-5197 |

| | | | | | | |
|----|------|------|--------|---|---------------------|-------------------|
| 28 | 16.2 | 70.4 | 34199 | 4 | US-09-902-540-1255 | Sequence 1255, Ap |
| 29 | 16.2 | 70.4 | 38119 | 4 | US-09-949-016-17503 | Sequence 17503, A |
| 30 | 16.2 | 70.4 | 52874 | 4 | US-09-949-016-14868 | Sequence 14868, A |
| 31 | 16.2 | 70.4 | 66247 | 4 | US-09-949-016-16009 | Sequence 16009, A |
| 32 | 16.2 | 70.4 | 86947 | 4 | US-09-949-016-11930 | Sequence 11930, A |
| 33 | 16.2 | 70.4 | 116966 | 4 | US-09-949-016-17557 | Sequence 17557, A |
| 34 | 16.2 | 70.4 | 162465 | 4 | US-09-949-016-14264 | Sequence 14264, A |
| 35 | 16.2 | 70.4 | 186959 | 4 | US-09-949-016-13125 | Sequence 13125, A |
| 36 | 16.2 | 70.4 | 325791 | 4 | US-09-768-185A-1 | Sequence 1, Appli |
| 37 | 16.2 | 70.4 | 767677 | 4 | US-09-949-016-12147 | Sequence 12147, A |
| 38 | 16.2 | 70.4 | 767677 | 4 | US-09-949-016-17361 | Sequence 17361, A |
| 39 | 16.2 | 70.4 | 786431 | 4 | US-09-751-389-3 | Sequence 3, Appli |
| 40 | 16 | 69.6 | 917 | 3 | US-09-221-017B-659 | Sequence 659, App |
| 41 | 15.8 | 68.7 | 601 | 4 | US-09-949-016-32251 | Sequence 32251, A |
| 42 | 15.8 | 68.7 | 601 | 4 | US-09-949-016-32252 | Sequence 32252, A |
| 43 | 15.8 | 68.7 | 601 | 4 | US-09-949-016-32253 | Sequence 32253, A |
| 44 | 15.8 | 68.7 | 601 | 4 | US-09-949-016-50387 | Sequence 50387, A |
| 45 | 15.8 | 68.7 | 601 | 4 | US-09-949-016-55234 | Sequence 55234, A |

ALIGNMENTS

RESULT 1
US-09-949-016-14890
; Sequence 14890, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14890
; LENGTH: 93364
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(93364)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14890
Query Match 81.7%; Score 18.8; DB 4; Length 93364;
Best Local Similarity 90.9%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 2 ATGTATTGATGGGATAGG 23
Db 48323 ATGTTTGGATGGGATAGG 48344
RESULT 2
US-09-949-016-13089
; Sequence 13089, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13089
; LENGTH: 89689
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(89689)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13089

Query Match      77.4%; Score 17.8; DB 4; Length 89689;
Best Local Similarity 90.5%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CATGTATTGATGGGATAGA 21
Db 14881 CATGTATTGATGGGACAGA 14901
|||||

RESULT 3
US-09-949-016-16179
; Sequence 16179, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16179
; LENGTH: 157644
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(157644)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16179

Query Match      77.4%; Score 17.8; DB 4; Length 157644;
Best Local Similarity 90.5%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTGATGGGATAGAG 22
Db 150480 ATGTATTGAGGGGGCTAGAG 150500
|||||

RESULT 4
US-09-949-016-16180
; Sequence 16180, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
```

```
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16180
; LENGTH: 157644
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(157644)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16180

Query Match      77.4%; Score 17.8; DB 4; Length 157644;
Best Local Similarity 90.5%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTGATGGGATAGAG 22
Db 150480 ATGTATTGAGGGGGCTAGAG 150500
|||||

RESULT 5
US-09-710-279-1277
; Sequence 1277, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1277
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-1277

Query Match      74.8%; Score 17.2; DB 4; Length 1026;
Best Local Similarity 86.4%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ATGTATTGATGGGATAGAG 23
Db 470 ATATATTAAATGGGATATAGG 491
|||||

RESULT 6
US-09-710-279-137
; Sequence 137, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 51291 TGTATTGGTGGGAGAG 51310

RESULT 11

US-09-313-294A-765
; Sequence 765, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 760
; SOFTWARE: PERL Program
; SEQ ID NO 765
; LENGTH: 209
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700549795H1
; NAME/KEY: unsure
; LOCATION: 29, 37, 39, 52, 57, 71, 76, 85, 118, 153-154, 201
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-765

Query Match 72.2%; Score 16.6; DB 4; Length 209;

Best Local Similarity 82.6%; Pred. No. 2.4e+02; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CATGTATTTGATGGGATAGG 23

Db 4 CATGTATCTGGTGGATGTGG 26

RESULT 12

US-09-513-999C-22576
; Sequence 22576, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 22576
; LENGTH: 211
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 40
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 184
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 202
; OTHER INFORMATION: k=g or t
US-09-513-999C-22576

Query Match 72.2%; Score 16.6; DB 4; Length 211;
Best Local Similarity 82.6%; Pred. No. 2.4e+02; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CATGTATTTGATGGGATAGG 23

Db 72 CATGAATTTGATGGGATATGG 94

RESULT 13

US-09-949-016-64747
; Sequence 64747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64747
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-64747

Query Match 72.2%; Score 16.6; DB 4; Length 601;

Best Local Similarity 82.6%; Pred. No. 2.8e+02; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CATGTATTTGATGGGATAGG 23

Db 189 CATGTGTGTGATGGTGAGAGG 211

RESULT 14

US-09-328-111-93
; Sequence 93, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 93
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-328-111-93

Query Match 72.2%; Score 16.6; DB 3; Length 627;
Best Local Similarity 82.6%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CATGTATTGATGGGATAGAGG 23
| | | | | | | | | | | | | | | | | | | | | |
Db 411 CATTTTGGATGGGATAGAGG 433

RESULT 15

US-09-221-017B-1071
; Sequence 1071, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 1071:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3880 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...3880

Query Match 72.2%; Score 16.6; DB 3; Length 3880;
Best Local Similarity 82.6%; Pred. No. 3.6e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 CATGTATTGATGGGATAGAGG 23
| | | | | | | | | | | | | | | | | | | | | |
Db 3191 CATGGATGTGATGGGATCGTGG 3213
Search completed: August 13, 2005, 06:48:44
Job time : 60.5213 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 03:26:33 ; Search time 1478.12 Seconds
(without alignments)
592.293 Million cell updates/sec

Title: US-10-673-854-1

Perfect score: 23

Sequence: 1 catgtatttgatgggtagagag 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

1: gb_est1.*

2: gb_est2.*

3: gb_hic.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_gss1.*

9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| C 1 | 18.8 | 81.7 | 394 | 1 | AI552179 |
| C 2 | 18.8 | 81.7 | 401 | 1 | AA445156 |
| C 3 | 18.8 | 81.7 | 570 | 8 | AZ662915 |
| C 4 | 18.8 | 81.7 | 616 | 5 | BQ829481 |
| C 5 | 18.8 | 81.7 | 659 | 9 | CE717779 |
| C 6 | 18.8 | 81.7 | 685 | 9 | CC514760 |
| C 7 | 18.8 | 81.7 | 946 | 2 | BF580964 |
| C 8 | 18.8 | 81.7 | 952 | 5 | BQ929148 |
| C 9 | 18.8 | 81.7 | 979 | 5 | BQ941451 |
| C 10 | 18.8 | 81.7 | 987 | 8 | CC204478 |
| C 11 | 18.8 | 81.7 | 989 | 4 | BG296600 |
| C 12 | 18.4 | 80.0 | 323 | 8 | BH854982 |
| C 13 | 18.4 | 80.0 | 365 | 7 | T75599 |
| C 14 | 18.4 | 80.0 | 372 | 5 | BP658944 |
| C 15 | 18.4 | 80.0 | 400 | 1 | AV814465 |
| C 16 | 18.4 | 80.0 | 404 | 8 | B24975 |
| C 17 | 18.4 | 80.0 | 406 | 1 | AV786094 |
| C 18 | 18.4 | 80.0 | 406 | 1 | AV806091 |
| C 19 | 18.4 | 80.0 | 409 | 5 | BP662526 |
| C 20 | 18.4 | 80.0 | 415 | 1 | AV786678 |
| C 21 | 18.4 | 80.0 | 452 | 6 | CD529428 |
| C 22 | 18.4 | 80.0 | 452 | 9 | CNS0052U |
| C 23 | 18.4 | 80.0 | 458 | 1 | AI928235 |
| C 24 | 18.4 | 80.0 | 483 | 1 | AA586206 |

| | | | | | | |
|------|------|------|------|---|-----------|---------------------|
| 25 | 18.4 | 80.0 | 525 | 9 | TA288A02P | AL487137 T. brucei |
| C 26 | 18.4 | 80.0 | 556 | 9 | TA59G11P | AL455675 T. brucei |
| 27 | 18.4 | 80.0 | 585 | 9 | TA189G12Q | AL476670 T. brucei |
| 28 | 18.4 | 80.0 | 908 | 7 | CK412987 | CK412987 AUF IpGill |
| 29 | 18.4 | 80.0 | 1214 | 3 | CNS0A6GM | BX824231 Arabidops |
| 30 | 18.4 | 80.0 | 1227 | 3 | CNS0A4TV | BX85041 Arabidops |
| 31 | 18.4 | 80.0 | 1271 | 3 | CNS0A77H | BX822669 Arabidops |
| 32 | 18.4 | 80.0 | 1293 | 3 | CNS0A6JM | BX823458 Arabidops |
| 33 | 18.4 | 80.0 | 1302 | 3 | CNS0A6IY | BX823867 Arabidops |
| 34 | 18.2 | 79.1 | 195 | 2 | BB584673 | BB584673 Arabidops |
| C 35 | 18.2 | 79.1 | 306 | 1 | AV204432 | AV204432 AV204432 |
| C 36 | 18.2 | 79.1 | 474 | 2 | BP869304 | BP869304 IL3-ET011 |
| C 37 | 18.2 | 79.1 | 538 | 8 | AQ143748 | AQ143748 HS_3075 B |
| 38 | 18.2 | 79.1 | 589 | 7 | CR791845 | CR791845 DKF2P468N |
| 39 | 18.2 | 79.1 | 611 | 7 | CK943116 | CK943116 4067003 B |
| 40 | 18.2 | 79.1 | 654 | 6 | CF073177 | CF073177 FE1 28 G0 |
| 41 | 18.2 | 79.1 | 674 | 8 | AQ040543 | AQ040543 CIT-HSP-2 |
| C 42 | 18.2 | 79.1 | 687 | 9 | CE737028 | CE737028 tigr-988- |
| C 43 | 18.2 | 79.1 | 713 | 8 | AZ246641 | AZ246641 RPCI-23-B |
| C 44 | 18.2 | 79.1 | 736 | 7 | CK943520 | CK943520 4067387 B |
| C 45 | 18.2 | 79.1 | 747 | 9 | CL839877 | CL839877 OR_Cba006 |

ALIGNMENTS

| | | | | | | |
|------------|---|--|------------|------|--------|-----------------|
| RESULT 1 | AI552179/c | AI552179 | 394 bp | mrna | linear | EST 23-MAR-1999 |
| LOCUS | VF58F07.Y1 | Soares mouse NBMH Mus musculus cDNA clone IMAGE:848005 | | | | |
| DEFINITION | 5' similar to gb:U07592 PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR BETA (HUMAN); mRNA sequence. | | | | | |
| ACCESSION | AI552179 | AI552179.1 | GI:4484553 | | | |
| VERSION | EST. | | | | | |
| KEYWORDS | Mus musculus | | | | | |
| SOURCE | Mus musculus (house mouse) | | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | | |
| REFERENCE | 1 (bases 1 to 394) | | | | | |
| AUTHORS | Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R., Waterston, R. and Wilson, R. | | | | | |
| TITLE | The WashU-NCI Mouse EST Project 1999 | | | | | |
| JOURNAL | Unpublished (1999) | | | | | |
| COMMENT | Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:500157 This read is a RESEQUENCE of a previously sequenced mouse clone correct orientation) Seq primer: -40RP from Gibco High quality sequence stop: 382. | | | | | |
| FEATURES | Location/Qualifiers | | | | | |
| source | 1..394 | | | | | |
| | /organism="Mus musculus" | | | | | |
| | /mol_type="mRNA" | | | | | |
| | /strain="C57BL/6J" | | | | | |
| | /db_xref="taxon:10090" | | | | | |
| | /clone="IMAGE:848005" | | | | | |
| | /sex="male" | | | | | |
| | /tissue_type="heart" | | | | | |
| | /dev_stage="4 weeks" | | | | | |
| | /lab_host="DH10B" | | | | | |
| | /clone_lib="Soares mouse NBMH" | | | | | |

adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
Query Match 81.7%; Score 18.8; DB 8; Length 570;
Best Local Similarity 90.9%; Pred. No. 7.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTTCATGGGATAGAGG 23
|||||
DB 502 ATGTATTTCATGGGATAGAGG 523

RESULT 4
BQ829481/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

BQ829481 616 bp mRNA linear EST 15-SEP-2002
LL6in20408 AFT024-subtracted library Mus musculus cDNA 5' similar
to peroxisome proliferator activated receptor delta, mRNA sequence.
BQ829481
BQ829481.1 GI:22861536
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 616)
Hackney, J.A., and Moore, K.A.
Lemischka, I.R., and Moore, K.A.
A molecular profile of a hematopoietic stem cell niche
Proc. Natl. Acad. Sci. U.S.A. 99 (20), 13061-13066 (2002)
122247628
12226475
Contact: Moore, Kateri A.
Department of Molecular Biology
Princeton University
217 Lewis Thomas Laboratory, Washington Road, Princeton, NJ 08544,
USA
Tel: 609 258 0605
Fax: 609 258 2759
Email: kmoore@molbio.princeton.edu
These ESTs are derived from a subtracted cDNA library enriched for
gene products expressed by a hematopoietic stem cell-supporting
stromal cell line, AFT024.
Seq primer: M13Reverse or T7.

FEATURES
source
1. .616
Location/Qualifiers

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/tissue_type="Fetal Liver"
/cell_type="Stromal cell"
/cell_line="AFT024"
/dev_stage="Embryonic day 14-14.5"
/lab_host="DH10B"
/clone_lib="AFT024-subtracted library"
/note="Organ: Fetal Liver; Vector: Sport 1; Site_1: Sal I;
Site_2: Not I; Two directionally cloned cDNA libraries
were made from a hematopoietic stem cell-supporting
stromal cell line (AFT024) and from a non-supporting
stromal cell line (2018). Subtractive hybridization was
performed by hybridization of the target, AFT024, single
stranded cDNA library in pSport1 to biotinylated RNA
transcribed from the driver, 2018 cDNA library in pSport2
with inserts cloned in the complementary orientation. The
AFT024-subtracted library contains 4.2x105 clones and is
depleted of common housekeeping gene products eg.
beta-actin and enriched for transcripts specific to
AFT024. For detailed protocols and additional information
please see our website at
<http://stromalcell.princeton.edu>."

ORIGIN

Query Match 81.7%; Score 18.8; DB 5; Length 616;
Best Local Similarity 90.9%; Pred. No. 7.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CATGTATTTCATGGGATAGAG 22
|||||
DB 284 CATGTCTTTGAAGGGATAGAG 263

RESULT 5
CE717779
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

CE717779 659 bp DNA linear GSS 29-SEP-2003
tigr-ges-dog-17000314586051 Dog Library Canis familiaris genomic,
genomic survey sequence.
CE717779
CE717779.1 GI:37037217
GSS.
Canis familiaris (dog)
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 659)
Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
22875432
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.

FEATURES
source
1. .659
Location/Qualifiers

/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN

Query Match 81.7%; Score 18.8; DB 9; Length 659;
Best Local Similarity 90.9%; Pred. No. 7.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTTCATGGGATAGAGG 23
|||||
DB 571 ATGTATTTCATGGGATAGAGG 592

RESULT 6
CC514760/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

CC514760 685 bp DNA linear GSS 17-JUN-2003
CH240_359D20.T7 CHORI-240 Bos taurus genomic clone CH240_359D20,
genomic survey sequence.
CC514760
CC514760.1 GI:31833048
GSS.
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 685)
Holt, R., Stott, J., Yang, G., Barber, S., Smalil, D., Prabhu, A.-L.,

Tsai, M., Cloutier, A., Lee, D., Gilm, N., Olson, T., Mayo, M., Butterfield, Y., Kirkpatrick, R., Liu, J., Guin, R., Chan, A., Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S., Schein, J., Marra, M., de Jong, P., McWilliam, S., Barris, W., Dalrymple, B. P. and Tellam, R.
Bovine BAC End Sequences from Library CHORI-240, PLATES 294 to 398
Unpublished (2003)
Other GSSs: CH240_359D20.TARBAC13P2
Contact: Rob Holt

Sequencing
The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6

Tel: 604-877-6085
Fax: 604-877-6276
Email: rholt@bcgsc.ca

Clones are derived from the bovine BAC library CHORI-240
(<http://www.chori.org/bacpac/bovine240.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(<http://www.chori.org/bacpac/ordering/information.htm>). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the British Columbia Genome Sciences Centre, Canada.
Plate: 359 row: D column: 20
Seq primer: T7
Class: BAC ends.

Location/Qualifiers

1. .685
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_359D20"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI; Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 81.7%; Score 18.8; DB 9; Length 685;
Best Local Similarity 90.9%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTTCATGGGATAGG 23

Db 149 AAGTATTTAATGGGATAGG 128

RESULT 7

BF580964/c
LOCUS 602100666F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4223944 5',
DEFINITION mRNA sequence.

ACCESSION BF580964

VERSION BF580964.1 GI:11654676

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 946)

/db_xref="taxon:10090"

/clone_lib="NCI_CGAP_Co24"

/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9813 row: c column: 17
High quality sequence stop: 699.

FEATURES

source

1. .946
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4223944"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 946;
Best Local Similarity 90.9%; Pred. No. 7.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CATGTATTTCATGGGATAGG 22

Db 224 CATGCTTTGAGGGATAGG 203

RESULT 8

BQ929148/c

LOCUS BQ929148

DEFINITION IMAGE:6477314 5', mRNA sequence.

ACCESSION BQ929148

VERSION BQ929148.1 GI:22344179

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM14019 row: b column: 03

High quality sequence stop: 288.

Location/Qualifiers

1. .952
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6477314"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 81.7%; Score 18.8; DB 5; Length 952;
Best Local Similarity 90.9%; Pred. No. 7.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 CATGTATTTCATGGGGATAGAG 22
Db 576 CATGCTCTTGAAGGGGATAGAG 555

RESULT 9
BQ941451/c
LOCUS
DEFINITION AGENCOURT 8819379 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6422948
5', mRNA sequence.
ACCESSION BQ941451
VERSION BQ941451.1 GI:22356929
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE BQ941451
AUTHORS AGENCOURT
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
cDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2600 row: 1 column: 21
High quality sequence start: 15
High quality sequence stop: 596.

FEATURES
source
1..979
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6422948"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_18"
/notes="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Query Match 81.7%; Score 18.8; DB 5; Length 979;
Best Local Similarity 90.9%; Pred. No. 7.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATGTATTTCATGGGGATAGAG 23
Db 864 ATGTATTTCATGGGGTATAGG 843

RESULT 10
CC204478/c
LOCUS
DEFINITION CC204478 CH261-131C12 Sp6.1 CH261 Gallus gallus genomic clone CH261-131C12,
genomic survey sequence.
ACCESSION CC204478
VERSION CC204478.1 GI:30486459
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 987)
Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E. and Wilson, R.
Gallus gallus BAC End Reads
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: Sp6 ATTAGTGACACTATAG
Class: BAC ends
High quality sequence start: 35
High quality sequence stop: 625.

FEATURES
source
1..987
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-131C12"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/notes="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - for library and clone
ordering information: http://www.chori.org/bacpac"

ORIGIN
Query Match 81.7%; Score 18.8; DB 8; Length 987;
Best Local Similarity 90.9%; Pred. No. 7.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CATGTATTTCATGGGGATAGAG 22
Db 371 CATGACTTTTCATGGGGATAGAG 350

RESULT 11
BG296600/c
LOCUS
DEFINITION 602393522F1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:4505236 5',
mRNA sequence.
ACCESSION BG296600
VERSION BG296600.1 GI:13059414
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 989)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10378 row: 1 column: 05
High quality sequence stop: 686.

FEATURES
source
1..989
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:4505236"

```

```

/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_94"
/notes="Organ: eye; Vector: pCMV-SORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 3.3 Kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN
Query Match      81.7%; Score 18.8; DB 4; Length 989;
Best Local Similarity 90.9%; Pred. No. 7.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CATGTTTGGGGGATAGAG 22
||||| ||||| ||||| ||||| |||||
Db 230 CATGCTTTGGAGGGGATAGAG 209

RESULT 12
BH854982/c
LOCUS
DEFINITION
  323 bp DNA linear GSS 08-JUL-2002
  SALK_087111.44.20.x Arabidopsis thaliana TDNA insertion lines
  Arabidopsis thaliana genomic clone SALK_087111.44.20.x, genomic
  survey sequence.
ACCESSION
  BH854982
VERSION
  BH854982.1 GI:21704572
KEYWORDS
  GSS.
SOURCE
  Arabidopsis thaliana (thale cress)
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
  1 (bases 1 to 323)
REFERENCE
  Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R.,
  Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
  Shinn,P., Zimmerman,J. and Ecker,J.R.
  A Sequence-Indexed Library of Insertion Mutations in the
  Arabidopsis Genome
  Unpublished (2001)
  Contact: Joseph R. Ecker
  Salk Institute Genomic Analysis Laboratory (SIGNAL)
  The Salk Institute for Biological Studies
  10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
  Tel: 858 453 4100 x1752
  Fax: 858 558 6379
  Email: ecker@salk.edu
  This is single pass sequence recovered from the left border of
  TDNA. This sequence lies within 300 bases of the 3' end of
  At3g02030 and an annotated exon of At3g02040.
  Class: TDNA tagged.
FEATURES
  source
  Location/Qualifiers
  1..323
    /organism="Arabidopsis thaliana"
    /mol_type="genomic DNA"
    /ecotype="Col-0"
    /db_xref="taxon:3702"
    /clone="SALK_087111.44.20.x"
    /clone_lib="Arabidopsis thaliana TDNA insertion lines"
    /notes="PCR was performed on Arabidopsis thaliana lines
    each of which contains one or more TDNA insertion
    elements. The resultant fragment for each line was
    directly sequenced to determine the genomic sequence at
    the site of insertion. Details of the protocols used can
    be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match      80.0%; Score 18.4; DB 8; Length 323;
Best Local Similarity 95.0%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GTATTTGATGGGATAGAG 23
||||| ||||| ||||| ||||| |||||
Db 81 GCATTTGATGGGATAGAG 62

RESULT 14
BP658944/c
LOCUS
DEFINITION
  372 bp mRNA linear EST 27-JUN-2004
  BP658944 RAFL19 Arabidopsis thaliana cDNA clone RAFL19-37-J11 3',
  mRNA sequence.
ACCESSION
  BP658944.1 GI:49310414
VERSION
  BP658944.1
KEYWORDS
  EST.
  Arabidopsis thaliana (thale cress)
  Arabidopsids

```


ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 372)
Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T.,
Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M.,
Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A. and Shinozaki, K.
Functional annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)
21932900
11910074

COMMENT

Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060

Email: mseki@rtc.riken.go.jp

reversed clone; Please visit our web site
(http://pfweb.gsc.riken.go.jp/) for further details.

FEATURES

source

1. .372
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL19-37-J11"
/tissue_type="mixture of silique and flower"
/lab_host="DH10B"
/clone_lib="RAFL19"
/notes="Site_1: BamHI; Site_2: SalI; Subtraction Library"

ORIGIN

Query Match 80.0%; Score 18.4; DB 5; Length 372;
Best Local Similarity 95.0%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GTATTGTGATGGGATAGAGG 23

DB 303 GCATTGTGATGGGATAGAGG 284

RESULT 15

AV814465/c

LOCUS

AV814465 400 bp mRNA linear EST 01-APR-2002
AV814465 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-83-B15 3',
mRNA sequence.

ACCESSION

VERSION

AV814465.1 GI:19856257

KEYWORDS

EST.

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 400)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)

REFERENCE

AUTHORS

Contact: Motoaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially

as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

and XhoI was ligated to modified Lambda FLX-1 vector (Carninci et

al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES

source

Location/Qualifiers
1. .400
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL09-83-B15"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/clone_lib="RAFL19"
/note="Site_1: BamHI; Site_2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"

ORIGIN

Query Match 80.0%; Score 18.4; DB 1; Length 400;

Best Local Similarity 95.0%; Pred. No. 1e+03;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GTATTGTGATGGGATAGAGG 23

DB 314 GCATTGTGATGGGATAGAGG 295

Search completed: August 13, 2005, 06:44:49

Job time : 1486.12 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 01:08:33 ; Search time 725.234 Seconds
(without alignments)
1536.704 Million cell updates/sec

Title: US-10-673-854-2

Perfect score: 23
Sequence: 1 cettttcatccaactaccactga 23

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| C 1 | 23 | 100.0 | 68726 | 9 AC016681 | AC016681 Homo sapi |
| C 2 | 23 | 100.0 | 100545 | 9 AL590379 | AL590379 Human DNA |
| C 3 | 23 | 100.0 | 129357 | 9 AL590492 | AL590492 Human DNA |
| C 4 | 21.4 | 93.0 | 59272 | 9 AL592490 | AL592490 Human DNA |
| C 5 | 19.4 | 84.3 | 93279 | 9 AL807757 | AL807757 Human DNA |
| C 6 | 18.8 | 81.7 | 155981 | 2 AC073361 | AC073361 Homo sapi |
| C 7 | 18.8 | 81.7 | 157122 | 2 AC102699 | AC102699 Mus muscu |
| C 8 | 18.8 | 81.7 | 165358 | 9 AL353782 | AL353782 Human DNA |
| C 9 | 18.8 | 81.7 | 166680 | 9 AC010132 | AC010132 Homo sapi |
| C 10 | 18.8 | 81.7 | 174558 | 2 CR382283 | CR382283 Danio rer |
| C 11 | 18.8 | 81.7 | 181241 | 9 AC080128 | AC080128 Homo sapi |
| C 12 | 18.8 | 81.7 | 184801 | 2 CR354436 | CR354436 Danio rer |
| C 13 | 18.8 | 81.7 | 187017 | 2 AC010101 | AC010101 Homo sapi |
| C 14 | 18.8 | 81.7 | 192318 | 9 AC011092 | AC011092 Homo sapi |
| C 15 | 18.8 | 81.7 | 193528 | 10 AL772234 | AL772234 Mouse DNA |
| C 16 | 18.8 | 81.7 | 204548 | 5 AC145960 | AC145960 Gallus ga |
| C 17 | 18.4 | 80.0 | 190996 | 10 AL671976 | AL671976 Mouse DNA |
| C 18 | 18.4 | 80.0 | 207077 | 2 AC097903 | AC097903 Rattus no |
| C 19 | 18.4 | 80.0 | 226926 | 10 AC122287 | AC122287 Mus muscu |

| | | | | | |
|------|------|------|--------|-------------|--------------------|
| C 20 | 18.4 | 80.0 | 267058 | 2 AC134185 | AC134185 Rattus no |
| C 21 | 18.2 | 79.1 | 549 | 5 AY323570 | AY323570 Euplectes |
| C 22 | 18.2 | 79.1 | 1041 | 5 AY329424 | AY329424 Muscicapa |
| C 23 | 18.2 | 79.1 | 1537 | 9 AF221594 | AF221594 Homo sapi |
| C 24 | 18.2 | 79.1 | 3001 | 6 BD265960 | BD265960 Schizophr |
| C 25 | 18.2 | 79.1 | 3001 | 6 AR244490 | AR244490 Sequence |
| C 26 | 18.2 | 79.1 | 30436 | 9 AC093016 | AC093016 Homo sapi |
| C 27 | 18.2 | 79.1 | 94847 | 2 AC091875 | AC091875 Homo sapi |
| C 28 | 18.2 | 79.1 | 101138 | 9 AC027006 | AC027006 Homo sapi |
| C 29 | 18.2 | 79.1 | 116983 | 8 OSJN00051 | AL606613 Oryza sat |
| C 30 | 18.2 | 79.1 | 117048 | 2 AC109629 | AC109629 Mus muscu |
| C 31 | 18.2 | 79.1 | 139552 | 10 AL731893 | AL731893 Mouse DNA |
| C 32 | 18.2 | 79.1 | 144865 | 9 CNS07EF7 | AL523558 Human chr |
| C 33 | 18.2 | 79.1 | 146313 | 8 OSJN00361 | AL731617 Oryza sat |
| C 34 | 18.2 | 79.1 | 151509 | 9 CNS05TE8 | AL358337 Human chr |
| C 35 | 18.2 | 79.1 | 166966 | 10 AC122015 | AC122015 Mus muscu |
| C 36 | 18.2 | 79.1 | 167030 | 2 AC025646 | AC025646 Homo sapi |
| C 37 | 18.2 | 79.1 | 174437 | 10 AC107781 | AC107781 Mus muscu |
| C 38 | 18.2 | 79.1 | 181541 | 9 AL136524 | AL136524 Human DNA |
| C 39 | 18.2 | 79.1 | 188755 | 9 AC013562 | AC013562 Homo sapi |
| C 40 | 18.2 | 79.1 | 189897 | 10 AC138769 | AC138769 Mus muscu |
| C 41 | 18.2 | 79.1 | 201508 | 2 AC026290 | AC026290 Homo sapi |
| C 42 | 18.2 | 79.1 | 213785 | 2 CR788255 | CR788255 Danio rer |
| C 43 | 18.2 | 79.1 | 225990 | 2 AC123627 | AC123627 Mus muscu |
| C 44 | 18.2 | 79.1 | 241913 | 2 AC094048 | AC094048 Rattus no |
| C 45 | 18.2 | 79.1 | 242318 | 10 AC097354 | AC097354 Mus muscu |

ALIGNMENTS

| | | | | | |
|------------|--|----------------------------|-----|--------|-----------------|
| RESULT 1 | AC016681/c | 68726 bp | DNA | linear | PRI 30-SEP-2000 |
| LOCUS | Homo sapiens BAC clone RP11-62H15 | from Y, complete sequence. | | | |
| DEFINITION | AC016681 | | | | |
| ACCESSION | AC016681.2 | GI:7321924 | | | |
| VERSION | HTG. | | | | |
| KEYWORDS | Homo sapiens (human) | | | | |
| SOURCE | Homo sapiens | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 68726) | | | | |
| AUTHORS | Sulston, J.E. and Waterston, R. | | | | |
| TITLE | Toward a complete human genome sequence | | | | |
| JOURNAL | Genome Res. 8 (11), 1097-1108 (1998) | | | | |
| MEDLINE | 99033792 | | | | |
| PUBMED | 9847074 | | | | |
| REFERENCE | 2 (bases 1 to 68726) | | | | |
| AUTHORS | Joshu, C., Stromatt, C. and Wedgeworth, P. | | | | |
| TITLE | The sequence of Homo sapiens BAC clone RP11-62H15 | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 3 (bases 1 to 68726) | | | | |
| AUTHORS | Waterston, R.H. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (04-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA | | | | |
| REFERENCE | 4 (bases 1 to 68726) | | | | |
| AUTHORS | Waterston, R.H. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (24-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA | | | | |
| REFERENCE | 5 (bases 1 to 68726) | | | | |
| AUTHORS | Waterston, R.H. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (04-APR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA | | | | |
| REFERENCE | 6 (bases 1 to 68726) | | | | |
| AUTHORS | Waterston, R.H. | | | | |
| TITLE | Direct Submission | | | | |

```

JOURNAL Submitted (07-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 7 (bases 1 to 68726)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Mar 24, 2000 this sequence version replaced gi:6524399.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The position of this clone was established as part of a
collaboration between the Human Chromosome Y Mapping Project
(Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen,
and David C. Page at the Whitehead Institute for Biomedical
Research, Cambridge MA) and the Washington University Genome
Sequencing Center, St. Louis MO.

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-362J16; the clone sequenced
to the right is RP11-218E11. Actual start of this clone is at base
position 1 of RP11-62H15.

FEATURES             Location/Qualifiers
     source            1..68726
                     /organism="Homo sapiens"
                     /mol_type="Genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="Y"
                     /map="Y"
                     /clone="RP11-62H15"
                     /clone_lib="RPCI-11"
     repeat_region    1..352
                     /rpt_family="L1"
     repeat_region    357..651
                     /rpt_family="Alu"
     repeat_region    676..769
                     /rpt_family="L1"
     repeat_region    1837..1873
                     /rpt_family="AT_rich"
     repeat_region    2274..2301
                     /rpt_family="AT_rich"
     repeat_region    2542..2572
                     /rpt_family="L1"

     repeat_region    3768..3973
                     /rpt_family="AT_rich"
     repeat_region    4258..4660
                     /rpt_family="MaLR"
     repeat_region    5234..5568
                     /rpt_family="MER1_type"
     repeat_region    6159..6713
                     /rpt_family="L1"
     repeat_region    6718..6943
                     /rpt_family="L1"
     repeat_region    6929..7694
                     /rpt_family="L1"
     repeat_region    7709..7887
                     /rpt_family=" (TA)n"
     repeat_region    7903..8247
                     /rpt_family="L1"
     repeat_region    8264..9031
                     /rpt_family="L1"
     repeat_region    9435..9583
                     /rpt_family="MIR"
     repeat_region    9584..9738
                     /rpt_family="MIR"
     repeat_region    10428..10482
                     /rpt_family="L2"
     repeat_region    10574..11996
                     /rpt_family="L1"
     repeat_region    11997..12024
                     /rpt_family="AT_rich"
     repeat_region    12040..13387
                     /rpt_family="L1"
     repeat_region    13390..13846
                     /rpt_family="L1"
     repeat_region    13837..14070
                     /rpt_family="L1"
     repeat_region    14074..15649
                     /rpt_family="L1"
     repeat_region    15648..16717
                     /rpt_family="L1"
     repeat_region    16720..16789
                     /rpt_family="A-rich"
     repeat_region    16827..17303
                     /rpt_family="L1"
     repeat_region    17417..17575
                     /rpt_family="L2"
     repeat_region    17725..17768
                     /rpt_family="AT_rich"
     repeat_region    18047..18132
                     /rpt_family="L1"
     repeat_region    18338..18648
                     /rpt_family="Alu"
     repeat_region    19586..19620
                     /rpt_family=" (CA)n"
     repeat_region    19704..19918
                     /rpt_family="MIR"
     repeat_region    19925..19980
                     /rpt_family=" (ATG)n"
     repeat_region    20305..20330
                     /rpt_family=" (CA)n"
     repeat_region    20333..20671
                     /rpt_family="MER2_type"
     repeat_region    21471..21557
                     /rpt_family="MaLR"
     repeat_region    21661..22655
                     /rpt_family="L1"
     repeat_region    22725..22971
                     /rpt_family="L1"
     repeat_region    23441..23462
                     /rpt_family="AT_rich"
     repeat_region    23982..24011
                     /rpt_family="AT_rich"
     repeat_region    24167..24258
                     /rpt_family="L1"

```

```

repeat_region 24399..24463 /rpt_family="AT_rich"
repeat_region 24468..24576 /rpt_family="L2"
repeat_region 24754..25021 /rpt_family="Retroviral"
repeat_region 25115..25575 /rpt_family="L2"
repeat_region 25715..26221 /rpt_family="MaLR"
repeat_region 26501..26523 /rpt_family="AT_rich"
repeat_region 26567..26925 /rpt_family="MaLR"

Query Match 100.0%; Score 23; DB 9; Length 68726;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTTTCATCCAACTACCACTGA 23
Db 57136 CCTTTTCATCCAACTACCACTGA 57114

```

```

RESULT 2
AL590379/c
LOCUS Human DNA sequence from clone RP11-343H6 on chromosome X, complete
sequence.
ACCESSION AL590379.7 GI:29466482
VERSION AL590379
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 100545)
AUTHORS Howden, P.
TITLE Direct Submission
JOURNAL Submitted (01-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 1, 2003 this sequence version replaced gi:289333281.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

RP11-343H6 is from the library RPCI-11.2 constructed by the group

of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6.

FEATURES

Location/Qualifiers
source 1..100545
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-343H6"
/clone_lib="RPCI-11.2"

ORIGIN

```

Query Match 100.0%; Score 23; DB 9; Length 100545;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTTTCATCCAACTACCACTGA 23
Db 67668 CCTTTTCATCCAACTACCACTGA 67646

```

RESULT 3

```

AL590492/c
LOCUS Human DNA sequence from clone RP11-88H5 on chromosome X, complete
sequence.
ACCESSION AL590492
VERSION AL590492.7 GI:15072594
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 129357)
AUTHORS Heath, P.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 1, 2001 this sequence version replaced gi:13929489.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

```

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

RP11-88H5 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-88H5 it may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone RP11-88H5 is at 1 in this sequence. The true left end of clone RP11-156J23 is at 129258 in this sequence.

FEATURES

Location/Qualifiers

as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

Location/Qualifiers
1. .93279
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-49014"
/clone_lib="RP11-11.1"

ORIGIN

Query Match 84.3%; Score 19.4; DB 9; Length 93279;
Best Local Similarity 95.2%; Pred. No. 90;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TTTTCATCCAACTACCACTGA 23
|||||
Db 19692 TTTTCATCCAACTACCACTGA 19672

RESULT 6

AC073361/c

LOCUS AC073361 155981 bp DNA linear HTG 08-JAN-2003
DEFINITION Homo sapiens chromosome 3 clone RP11-160F1, WORKING DRAFT SEQUENCE,
20 unordered pieces.

ACCESSION

AC073361.8 GI:20335704

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 155981)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayete,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroil,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louleaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Mahehwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S., Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojlobokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshkari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.P., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
Unpublished
2 (bases 1 to 155981)
Worley,K.C.
Direct Submission
Submitted (15-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 155981)
Worley,K.C.
Direct Submission
Submitted (08-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:15887171.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Drafting Center Code: BCM

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HBQ

Center clone name: RP11-160F1

----- Summary Statistics

Sequencing vector: M13;

Chemistry: Dye-terminator Big Dye; 40% of reads

Chemistry: Dye-terminator Big Dye; 60% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 141831 bases at least Q40

Consensus quality: 146281 bases at least Q30

Consensus quality: 148168 bases at least Q20

Estimated insert size: 155230; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 20 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 2542: contig of 2542 bp in length

* 2543 2642: gap of unknown length

* 2643 4841: contig of 2199 bp in length

* 4842 4941: gap of unknown length

* 4942 7369: contig of 2428 bp in length

* 7370 7469: gap of unknown length

* 7470 11527: contig of 4058 bp in length

* 11528 11627: gap of unknown length

* 11628 16423: contig of 4796 bp in length

* 16424 16523: gap of unknown length

* 16524 20209: contig of 3686 bp in length

* 20210 20309: gap of unknown length

* 20310 24391: contig of 3982 bp in length

* 24392 24991: gap of unknown length

* 24992 27953: contig of 3562 bp in length

* 27954 28053: gap of unknown length

* 28054 32347: contig of 4194 bp in length

* 32348 37027: contig of 4680 bp in length

* 37028 37127: gap of unknown length

* 37128 41991: contig of 4864 bp in length

* 41992 42091: gap of unknown length

* 42092 47187: contig of 5095 bp in length

* 47188 47286: gap of unknown length

* 47287 53916: contig of 6630 bp in length

* 47287

REFERENCE
AUTHORS

* 53917 54016: gap of unknown length
 * 54017 59973: contig of 5957 bp in length
 * 59974 60073: gap of unknown length
 * 60074 67148: contig of 7075 bp in length
 * 67149 67248: gap of unknown length
 * 67249 77842: contig of 10594 bp in length
 * 77843 77942: gap of unknown length
 * 77943 91209: contig of 13267 bp in length
 * 91210 91309: gap of unknown length
 * 91310 104946: contig of 13637 bp in length
 * 104947 105046: gap of unknown length
 * 105047 126327: contig of 21280 bp in length
 * 126327 126427: gap of unknown length
 * 126427 155981: contig of 29555 bp in length.

FEATURES

Location/Qualifiers
 1. .155981
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-160P1"

ORIGIN

Query Match 81.78; Score 18.8; DB 2; Length 155981;

Best Local Similarity 90.94; Pred. No. 1.7e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CTTTTCATCAACTTACCACCTGA 23

Db 143609 CGTTTCATCCAGCTACCACCTGA 143588

RESULT 7

AC102699/c

LOCUS AC102699 157122 bp DNA linear HTG 03-MAR-2003
 DEFINITION Mus musculus clone RP24-268F1, WORKING DRAFT SEQUENCE, 6 unordered
 pieces.

ACCESSION AC102699 GI:28631330

VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Birren.B., Nusbaum.C. and Lander.E.

1 (bases 1 to 157122)

Mus musculus, clone RP24-268F1

Unpublished

2 (bases 1 to 157122)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
 Lamazares,R., Landers,T., Lehotsky,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome

TITLE

JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 157122)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Faro,S.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Gardyna,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramaamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo.A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Genome Center
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L19308

Center clone name: 268_F_1

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 156083 bases at least Q40

Consensus quality: 156406 bases at least Q30

Consensus quality: 156521 bases at least Q20

Insert size: 151000; agarose-fp

Insert size: 156622; sum-of-contigs

Quality coverage: 10.5 in Q20 bases; agarose-fp

Quality coverage: 10.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 51217: contig of 51217 bp in length
 * 51218 51317: gap of 100 bp
 * 51318 52366: contig of 1049 bp in length
 * 52367 52466: gap of 100 bp
 * 52467 114294: contig of 61828 bp in length
 * 114295 114394: gap of 100 bp
 * 114395 121686: contig of 7292 bp in length
 * 121687 121786: gap of 100 bp
 * 121787 153936: contig of 32150 bp in length
 * 153937 154036: gap of 100 bp
 * 154037 157122: contig of 3086 bp in length.

Location/Qualifiers

1. .157122

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

FEATURES
 source

```

/misc_feature      /clone="RP24-268P1"
                  /clone_lib="RPC1-24 Male Mouse BAC"
                  1. .51217
                  /note="assembly_fragment
clone end:SP6
vector_side:left"
51318_52366
/misc_feature      /note="assembly_fragment"
52467_114294
/misc_feature      /note="assembly_fragment"
114395_121686
/misc_feature      /note="assembly_fragment"
121787_153936
/misc_feature      /note="assembly_fragment"
154037_157122
/misc_feature      /note="assembly_fragment
clone end:T7
vector_side:right"

ORIGIN
Query Match      81.7%; Score 18.8; DB 2; Length 157122;
Best Local Similarity 90.9%; Pred. NO. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2      CTTTTCATCCAACTACCACTGA 23
|||||
Db      57866 CTTTTCATCCAACTAGAACTGA 57845

RESULT 8
AL353782/c
LOCUS      AL353782
DEFINITION      Human DNA sequence from clone RP11-207C11 on chromosome 9, complete
sequence.
ACCESSION      AL353782
VERSION      AL353782.23 GI:18476550
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sehra,H.
Direct Submission
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:17127880.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30) ; an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Swi,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
Chromosome 9, constructed by the Sanger Centre
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-207C11 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm

```

```

VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-207C11 it may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP11-207C11 is at 165358 in this
sequence. The true right end of clone RP11-349P17 is at 2000 in
this sequence.
FEATURES
             location/Qualifiers
             source
             1. .165358
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="9"
             /clone="RP11-207C11"
             /clone_lib="RPC1-11.1"
             /clone="RP11-207C11"
             /note="Single clone region. Assembly confirmed by
restriction digest data."
             114473..114543
             /note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
             misc_feature
             60739..60742
             /note="Single clone region. Assembly confirmed by
restriction digest data."
             114473..114543
             misc_feature
             114473..114543

ORIGIN
Query Match      81.7%; Score 18.8; DB 9; Length 165358;
Best Local Similarity 90.9%; Pred. NO. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2      CTTTTCATCCAACTACCACTGA 23
|||||
Db      129391 CTTTTCATCCAACTCCCACTGA 129370

RESULT 9
AC010132/c
LOCUS      AC010132
DEFINITION      Homo sapiens BAC clone RP11-111K18 from 7p11.2-p2, complete
sequence.
ACCESSION      AC010132
VERSION      AC010132.5 GI:10440742
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston,J.E. and Waterston,R.
1 (bases 1 to 166680)
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED      9847074
REFERENCE      2 (bases 1 to 166680)
AUTHORS      Andrews,S., Wohldmann,P. and Scherger,E.
TITLE      The sequence of Homo sapiens BAC clone RP11-111K18
JOURNAL      Unpublished
MEDLINE
PUBMED
REFERENCE      2 (bases 1 to 166680)
AUTHORS      Andrews,S., Wohldmann,P. and Scherger,E.
TITLE      The sequence of Homo sapiens BAC clone RP11-111K18
JOURNAL      Unpublished
REFERENCE      3 (bases 1 to 166680)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (13-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 166680)
AUTHORS      Waterston,R.
TITLE      Direct Submission
JOURNAL      Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 30, 2000 this sequence version replaced gi:8954217.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics

```



```

repeat_region 24091..25047
/rpt_family="L2"
repeat_region 25091..25312
/rpt_family="L2"
repeat_region 25375..25549
/rpt_family="L2"
repeat_region 25663..25817
/rpt_family="MER1_type"
repeat_region 27098..27212
/rpt_family="L2"
repeat_region 27657..27925
/rpt_family="MaLR"
repeat_region 28020..28061
/rpt_family="L2"
repeat_region 28527..28780
/rpt_family="Alu"

Query Match 81.7%; Score 18.8; DB 9; Length 166680;
Best Local Similarity 90.9%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CTTTTCATCCAACTACCACTGA 23
| ||||| ||||| ||||| |||||
Db 71916 CATTTCATCCAACTACCACTG 71895

RESULT 10
CR382283/c
LOCUS 174558 bp DNA linear HTG 30-MAR-2004
DEFINITION Danio rerio clone DKYEP-32G11, *** SEQUENCING IN PROGRESS ***, 9
unordered pieces.
ACCESSION CR382283
VERSION 2.2 GI:46019446
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
Sims, S.
Direct Submission
Submitted (29-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 1, 2004 this sequence version replaced gi:45774318.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfsh-help@sanger.ac.uk
----- Project Information
Center project name: zKp32G11
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 171524 bases at least Q40
Consensus quality: 172036 bases at least Q30
Consensus quality: 172486 bases at least Q20
Insert size: 173758; sum-of-contigs
Insert size: 172615; 6.9% error; agarose-fp
Quality coverage: 7.43x in Q20 bases; sum-of-contigs Quality
coverage: 7.76x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8310: contig of 8310 bp in length

repeat_region 8410: gap of 100 bp
* 8311 52211: contig of 43801 bp in length
* 8411 52311: gap of 100 bp
* 52312 71637: contig of 19326 bp in length
* 71638 71737: gap of 100 bp
* 71738 79714: contig of 7977 bp in length
* 79715 79814: gap of 100 bp
* 79815 107925: contig of 28111 bp in length
* 107926 108025: gap of 100 bp
* 108026 139165: contig of 31140 bp in length
* 139166 139265: gap of 100 bp
* 139266 142778: contig of 3513 bp in length
* 142779 142878: gap of 100 bp
* 142879 158385: contig of 15507 bp in length
* 158386 158486: gap of 100 bp
* 158487 174558: contig of 16073 bp in length.

FEATURES
Location/Qualifiers
source
1..174558
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKYEP-32G11"
/clone_lib="DanioKeyPilot"
1..8310
/note="assembly fragment:00142
fragment_chain:1"
misc_feature
8411..52211
/note="assembly fragment:01851
fragment_chain:1"
misc_feature
52312..71637
/note="assembly fragment:00673
fragment_chain:1"
misc_feature
71738..79714
/note="assembly fragment:00067
fragment_chain:1"
misc_feature
79815..107925
/note="assembly fragment:01363
fragment_chain:1"
misc_feature
108026..139165
/note="assembly fragment:00960
fragment_chain:1"
misc_feature
139266..142778
/note="assembly fragment:00035.0"
misc_feature
142879..158385
/note="assembly fragment:00253"
misc_feature
158486..174558
/note="assembly fragment:00432"

ORIGIN
Query Match 81.7%; Score 18.8; DB 2; Length 174558;
Best Local Similarity 90.9%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTTCATCCAACTACCACTG 22
| ||||| ||||| ||||| |||||
Db 78871 CCTTTATCCACTACCACTG 78850

RESULT 11
AC080128
LOCUS Homo sapiens 181241 bp DNA linear PRI 30-JUL-2002
DEFINITION Homo sapiens 3 BAC RP11-404G16 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
ACCESSION AC080128
VERSION AC080128.21 GI:21591797
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 181241)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,

```

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buay, C., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honsi, P., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, J., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, D.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Monabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubo, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabors, P., Tamerisa, K., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 181241)
Worley, K.C.

Direct Submission
Submitted (25-SEP-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 181241)
Worley, K.C.

Direct Submission
Submitted (26-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 181241)
Worley, K.C.

Direct Submission
Submitted (26-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 181241)
Worley, K.C.

Direct Submission
Submitted (30-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 26, 2002 this sequence version replaced gi:21206028.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

| source | Location/Qualifiers |
|---------------|---------------------------|
| repeat_region | 1. .181241 |
| repeat_region | /organism="Homo sapiens" |
| repeat_region | /mol_type="genomic DNA" |
| repeat_region | /db_xref="taxon:9606" |
| repeat_region | /chromosome="3" |
| repeat_region | /clone="RP11-404G16" |
| repeat_region | complement(214. .380) |
| repeat_region | /rpt_family="MER5A" |
| repeat_region | complement(5542. .5592) |
| repeat_region | /rpt_family="L2" |
| repeat_region | 6194. .6294 |
| repeat_region | /standard_name="WI-11301" |
| repeat_region | complement(6410. .6721) |
| repeat_region | /rpt_family="AluSx" |
| repeat_region | complement(7148. .7237) |
| repeat_region | /rpt_family="MER5A" |
| repeat_region | complement(7965. .8133) |
| repeat_region | /rpt_family="MIR" |
| repeat_region | complement(8977. .9063) |
| repeat_region | /rpt_family="MLT1I" |
| repeat_region | complement(10099. .10247) |
| repeat_region | /rpt_family="L2" |
| repeat_region | complement(11065. .11188) |
| repeat_region | /rpt_family="MIR" |
| repeat_region | 11305. .11335 |
| repeat_region | /rpt_family="GC_rich" |
| repeat_region | 11449. .11490 |
| repeat_region | /rpt_family="GC_rich" |
| repeat_region | complement(12269. .12403) |
| repeat_region | /rpt_family="L2" |
| repeat_region | complement(12870. .13115) |
| repeat_region | /rpt_family="MER102" |
| repeat_region | 13127. .13156 |
| repeat_region | /rpt_family="(CA)n" |
| repeat_region | 13343. .13451 |
| repeat_region | /rpt_family="L2" |
| repeat_region | complement(13490. .14235) |
| repeat_region | /rpt_family="LIME1" |
| repeat_region | complement(14245. .14279) |
| repeat_region | /rpt_family="LIME2" |
| repeat_region | complement(14280. .14584) |
| repeat_region | /rpt_family="AluSx" |

```

repeat_region complement(14585..14886)
repeat_region /rpt_family="LIME2"
repeat_region complement(15313..15378)
repeat_region /rpt_family="LIMD2"
repeat_region 15379..15691
repeat_region /rpt_family="AluSq"
repeat_region complement(15692..15955)
repeat_region /rpt_family="LIMD2"
repeat_region complement(15956..16258)
repeat_region /rpt_family="AluSx"
repeat_region complement(16259..16359)
repeat_region /rpt_family="LIMD2"
repeat_region 16442..16603
repeat_region /rpt_family="LIME2"
repeat_region complement(16640..16811)
repeat_region /rpt_family="LIMD1"
repeat_region complement(16812..17088)
repeat_region /rpt_family="LIPA7"
repeat_region complement(17089..17745)
repeat_region /rpt_family="LIMD1"
repeat_region complement(17752..17990)
repeat_region /rpt_family="LIME"
repeat_region complement(18007..18439)
repeat_region /rpt_family="LIMEc"
repeat_region 18446..18493
repeat_region /rpt_family="AT rich"
repeat_region complement(18532..19199)
repeat_region /rpt_family="LIMEc"
repeat_region complement(19437..19934)
repeat_region /rpt_family="LIMCc"
repeat_region 19974..20595
repeat_region /rpt_family="LIME2"
repeat_region 20632..20968
repeat_region /rpt_family="LIME2"
STS 20875..20981
repeat_region /standard_name="D3S2333"
repeat_region 20969..21275
repeat_region /rpt_family="AluSx"
repeat_region 21276..21288
repeat_region /rpt_family="LIME2"
repeat_region complement(21921..22221)

Query Match 81.7%; Score 18.8; DB 9; Length 181241;
Best Local Similarity 90.9%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CTTTTCATCCAACTACCACCTGA 23
Db 106165 CGTTTCATCCAGCTACCACCTGA 106186

RESULT 12
CR354436 184801 bp DNA linear HTG 10-OCT-2004
LOCUS Danio rerio clone DKEY-269F18, WORKING DRAFT SEQUENCE, 4 unordered
DEFINITION Pieces.
ACCESSION CR354436
VERSION CR354436.8 GI:54021885
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 184801)
AUTHORS Almeida, J.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Oct 10, 2004 this sequence version replaced gi:53142909.
----- Genome Center
Center: Wellcome Trust Sanger Institute

```

```

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
----- Project Information
Center project name: zK269F18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 184321 bases at least Q40
Consensus quality: 184412 bases at least Q30
Consensus quality: 184448 bases at least Q20
Insert size: 184501; sum-of-contigs
Insert size: 174924; 5.4% error; agarose-fp
Quality coverage: 9.52x in Q20 bases; sum-of-contigs Quality
coverage: 10.35x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 154052: contig of 154052 bp in length
* 154053 154152: gap of 100 bp
* 154153 161873: contig of 7721 bp in length
* 161874 161973: gap of 100 bp
* 161974 181767: contig of 19794 bp in length
* 181768 181867: gap of 100 bp
* 181868 184801: contig of 2934 bp in length.
* 181868 184801: contig of 2934 bp in length.
* Location/Qualifiers
* 1..184801
* /organism="Danio rerio"
* /mol_type="genomic DNA"
* /db_xref="taxon:7955"
* /clone="DKEY-269F18"
* /clone_lib="Daniokey"
* 1..154052
* /note="assembly fragment:02802"
* fragment_chain:1
* 154153..161873
* /note="assembly fragment:01702"
* fragment_chain:1
* 161974..181767
* /note="assembly fragment:01449"
* fragment_chain:1
* 181868..184801
* /note="assembly fragment:02662"

ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 184801;
Best Local Similarity 90.9%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTTTCATCCAACTACCACCTG 22
Db 31456 CGTTTCATCCAGCTACCACCTG 31477

RESULT 13
AC010101/c 187017 bp DNA linear HTG 08-AUG-2000
LOCUS Homo sapiens chromosome 7 clone RP11-447C21, WORKING DRAFT
DEFINITION SEQUENCE, 8 unordered pieces.
ACCESSION AC010101
VERSION AC010101.8 GI:9739351
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 187017)

```


Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (12-JAN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jan 2, 2002 this sequence version replaced gi:12718900.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1460
 Center clone name: 1_H15

FEATURES

Location/Qualifiers

source

1..192318
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-1H15"
 /clone_lib="RP11-1 Human Male BAC"
 complement(1..1561)
 /rpt_family="L1PA12"
 1560..2087
 /rpt_family="L1P"
 2199..2257
 /rpt_family="T-rich"
 complement(2258..2620)
 /rpt_family="L1PB3"
 2621..2668
 /rpt_family="(TG)n"
 complement(2669..2993)
 /rpt_family="L1PB3"
 2994..3293
 /rpt_family="AluSx"
 complement(3294..3693)
 /rpt_family="L1PB3"
 complement(4255..4469)
 /rpt_family="AluJb"
 complement(4471..4583)
 /rpt_family="FRAM"
 4895..5153
 /rpt_family="L1ME4A"
 5163..5192
 /rpt_family="(CATA)n"
 5337..5449
 /rpt_family="MIR"
 7233..7484
 /rpt_family="Tigger7"
 7471..7868
 /rpt_family="Tigger7"
 complement(8459..8574)
 /rpt_family="Charlie1"
 complement(8575..8868)
 /rpt_family="AluSq"
 complement(8869..9645)
 /rpt_family="Charlie1"
 complement(9662..9817)
 /rpt_family="Charlie1"
 complement(9815..9877)
 /rpt_family="Charlie1"
 complement(9979..10110)
 /rpt_family="MER8"
 complement(10639..10704)
 /rpt_family="MIR"
 complement(10709..11976)
 /rpt_family="L1M4"
 11977..12296
 /rpt_family="MER2"

repeat_region

complement(12297..12551)
 /rpt_family="L1M4"

repeat_region

complement(12556..12726)
 /rpt_family="LTR66"

repeat_region

12727..13026
 /rpt_family="AluSx"

repeat_region

complement(13027..13237)
 /rpt_family="LTR66"

repeat_region

complement(13227..13369)
 /rpt_family="LTR10E"

repeat_region

complement(13576..13610)
 /rpt_family="LTR66"

repeat_region

complement(13612..13883)
 /rpt_family="LTR66"

repeat_region

complement(13884..14319)
 /rpt_family="L1M4"

repeat_region

14320..14614
 /rpt_family="AluJo"

repeat_region

complement(14615..14777)
 /rpt_family="L1M4"

repeat_region

complement(17258..18526)
 /rpt_family="L1MC4"

repeat_region

19291..19340
 /rpt_family="AT-rich"

repeat_region

complement(19811..19952)
 /rpt_family="MIR"

repeat_region

21462..21628
 /rpt_family="MIR3"

repeat_region

complement(21773..22919)
 /rpt_family="L1PA7"

repeat_region

22920..23164
 /rpt_family="AluY"

repeat_region

24503..24529
 /rpt_family="(TG)n"

repeat_region

24694..24716
 /rpt_family="AT-rich"

repeat_region

complement(25358..25824)
 /rpt_family="L2"

repeat_region

complement(25825..25896)
 /rpt_family="L2"

repeat_region

complement(26035..26119)
 /rpt_family="L1MB4"

repeat_region

26154..26437
 /rpt_family="AluJo"

repeat_region

26895..26923
 /rpt_family="AT-rich"

repeat_region

29149..29275
 /rpt_family="CT-rich"

repeat_region

29569..29589
 /rpt_family="AT-rich"

repeat_region

29983..30238
 /rpt_family="AluJo"

repeat_region

30969..31229
 /rpt_family="L1MA8"

repeat_region

complement(31448..31744)
 /rpt_family="AluJo"

repeat_region

32030..32085
 /rpt_family="MER20B"

repeat_region

32295..32598
 /rpt_family="AluSx"

repeat_region

32970..33221
 /rpt_family="L1ME3"

repeat_region

33248..33473
 /rpt_family="L1M3e"

repeat_region

33760..33854
 /rpt_family="L1ME1"

repeat_region

33866..34308
 /rpt_family="L1ME1"

repeat_region

34318..34813
 /rpt_family="L1MB7"

repeat_region

34826..35443
 /rpt_family="L1MCS"

repeat_region

35722..35763

QY 2 CTTTTCATCCAACTACCACTGA 23
|||||
Db 54176 CTTTTCATCCAACTACCACTGA 54155
Search completed: August 13, 2005, 05:04:15
Job time : 732.234 secs

repeat_region /rpt_family="AT rich"
complement(35814..35916)
Query Match 81.7%; Score 18.8; DB 9; Length 192318;
Best Local Similarity 90.9%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CTTTTCATCCAACTACCACTGA 23
|||||
Db 72240 CATTTCATCCAACTACCACTGA 72219

RESULT 15
AL772234/c
LOCUS
DEFINITION Mouse DNA sequence from clone RP23-413M22 on chromosome 11,
complete sequence.
ACCESSION AL772234
VERSION AL772234.5 GI:23476661
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Pearce, A.
Direct Submission
Submitted (01-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 2, 2002 this sequence version replaced gi:23337380.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e. phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-413M22 is
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6.
Location/Qualifiers
1..193528
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-413M22"
/clone_lib="RPCI-23"

FEATURES
source
1..193528
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-413M22"
/clone_lib="RPCI-23"

ORIGIN
Query Match 81.7%; Score 18.8; DB 10; Length 193528;
Best Local Similarity 90.9%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 12, 2005, 22:21:57 ; Search time 184.245 Seconds
(without alignments)
738.985 Million cell updates/sec

Title: US-10-673-854-2

Perfect score: 23

Sequence: 1 ccttttcacactaccactga 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------|--------------------|
| C 1 | 18.2 | 79.1 | 3001 | 3 AAHS1779 | Aahs1779 Chromosom |
| C 2 | 18.2 | 79.1 | 53242 | 11 ACN43926 | Acn43926 Human gen |
| C 3 | 17.8 | 77.4 | 110000 | 12 AD079173_0 | Ado79173 KLF12 gen |
| C 4 | 17.4 | 75.7 | 50287 | 11 ACN44848 | Acn44848 Mouse gen |
| C 5 | 17.4 | 75.7 | 81684 | 13 ABD33502 | Abd33502 Murine ca |
| C 6 | 17.2 | 74.8 | 377 | 3 AAC28339 | Aac28339 Human sec |
| C 7 | 17.2 | 74.8 | 439 | 8 ABX38985 | Abx38985 Bovine ES |
| C 8 | 17.2 | 74.8 | 447 | 13 ADS59130 | Ads59130 Bacterial |
| C 9 | 17.2 | 74.8 | 512 | 8 ABZ36846 | Abz36846 Human GEN |
| C 10 | 17.2 | 74.8 | 588 | 12 ACH76108 | Ach76108 Human gen |
| C 11 | 17.2 | 74.8 | 661 | 6 ABQ99614 | Abq99614 Human mem |
| C 12 | 17.2 | 74.8 | 728 | 3 AAC93484 | Aac93484 Human sec |
| C 13 | 17.2 | 74.8 | 728 | 8 ABZ73565 | Abz73565 Secretd |
| C 14 | 17.2 | 74.8 | 728 | 8 ADA98080 | Ada98080 Human sec |
| C 15 | 17.2 | 74.8 | 728 | 10 ABZ67162 | Abz67162 Human sec |
| C 16 | 17.2 | 74.8 | 807 | 12 ACH89809 | Ach89809 Human gen |
| C 17 | 17.2 | 74.8 | 1107 | 5 AAS70101 | Aas70101 DNA encod |
| C 18 | 17.2 | 74.8 | 1164 | 9 AAL56877 | Aal56877 DNA encod |
| C 19 | 17.2 | 74.8 | 1194 | 9 AAL56878 | Aal56878 Human ade |
| C 20 | 17.2 | 74.8 | 1274 | 8 ABZ74428 | Abz74428 Secretd |

| | | | | | |
|------|------|------|--------|-------------|---------------------|
| 21 | 17.2 | 74.8 | 1274 | 8 ADA98866 | Ada98866 Human sec |
| 22 | 17.2 | 74.8 | 1374 | 10 ABZ67985 | Abz67985 Human sec |
| 23 | 17.2 | 74.8 | 1402 | 2 AAQ12228 | Aaq12228 Aq41 shor |
| 24 | 17.2 | 74.8 | 1402 | 2 AAQ24432 | Aaq24432 Tak short |
| 25 | 17.2 | 74.8 | 1939 | 2 AAQ12227 | Aaq12227 Aq41 long |
| 26 | 17.2 | 74.8 | 1939 | 2 AAQ24431 | Aaq24431 Tak long |
| C 27 | 17.2 | 74.8 | 2061 | 8 ABX71138 | Abx71138 Novel hum |
| C 28 | 17.2 | 74.8 | 3076 | 2 AAV73003 | Aav73003 Human adu |
| C 29 | 17.2 | 74.8 | 3076 | 6 ABQ92028 | Abq92028 Human pol |
| C 30 | 17.2 | 74.8 | 3544 | 6 ABA01937 | Abao1937 Human qui |
| C 31 | 17.2 | 74.8 | 4731 | 8 ACA45015 | ACA45015 Prokaryot |
| C 32 | 17.2 | 74.8 | 4770 | 10 ADF00541 | Adf00541 Bacterial |
| C 33 | 17.2 | 74.8 | 11866 | 1 AAN40141 | Aan40141 Sequence |
| C 34 | 17.2 | 74.8 | 11873 | 1 AAN40176 | Aan40176 Sequence |
| C 35 | 17.2 | 74.8 | 38736 | 6 ABQ99652 | Abq99652 Human mem |
| C 36 | 17.2 | 74.8 | 57728 | 4 AAC87588 | Aac87588 Human 9pl |
| C 37 | 17.2 | 74.8 | 162450 | 3 AAZ86967 | Aaz86967 Retinobla |
| C 38 | 16.8 | 73.0 | 128 | 4 AAI27140 | Aai27140 Probe #17 |
| C 39 | 16.8 | 73.0 | 128 | 4 ABA75421 | Aba75421 Human foe |
| C 40 | 16.8 | 73.0 | 128 | 4 AAI55992 | Aai55992 Probe #24 |
| C 41 | 16.8 | 73.0 | 128 | 4 ABA40051 | Abao40051 Probe #18 |
| C 42 | 16.8 | 73.0 | 128 | 4 AAK50052 | Aak50052 Human bon |
| C 43 | 16.8 | 73.0 | 128 | 4 AAK23986 | Aak23986 Human bra |
| C 44 | 16.8 | 73.0 | 128 | 4 ABS49694 | Abs49694 Human liv |
| C 45 | 16.8 | 73.0 | 396 | 5 AAH65658 | Aah65658 C glutami |

ALIGNMENTS

RESULT 1

AAHS1779/c

ID AAHS1779 standard; DNA; 3001 BP.

XX AC AAHS1779;

XX DT 29-AUG-2001 (first entry)

XX DE Chromosome 13q31-q33 biallelic marker containing amplicon SEQ ID 191.

XX KW sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;
XX KW biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.
XX OS Homo sapiens.
XX PN WO2000S8510-A2.
XX PD 05-OCT-2000.
XX PF 30-MAR-2000; 2000WO-IB000435.
XX PR 30-MAR-1999; 99US-0126903P.
XX PR 30-APR-1999; 99US-0131971P.
XX PR 30-APR-1999; 99US-0132065P.
XX PR 14-JUL-1999; 99US-0143928P.
XX PR 27-JUL-1999; 99US-0145915P.
XX PR 29-JUL-1999; 99US-0146452P.
XX PR 29-JUL-1999; 99US-0146453P.
XX PR 28-OCT-1999; 99US-0162288P.
(GEST) GENSET.
Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihain B;
Bessieux L;
WPI; 2000-619082/59.
Polynucleotides comprising sequences from sbg1 and g35018 biallelic
markers are used for genotyping and detecting schizophrenia or bipolar
disorder and predisposition to these disorders.
Claim 2; Page 684-685; 737pp; English.

CC AAH51601 represents a human genomic nucleotide sequence comprising sbg1,
CC g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the
CC human chromosome 13q31-q33 locus. The nucleotide sequences contain
CC biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and
CC AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein
CC products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018
CC cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used
CC to isolate sbg1 cDNAs, while sbg1 exons from different primates are
CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of
CC amplicons which comprise biallelic markers located on the chromosome
CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Biallelic markers
CC are represented in the sequences by degenerate/undefined base codes. PCR
CC primers AAH51818 and AAH51819 are used in the isolation of sequences of
CC the invention. The biallelic marker containing nucleotide sequences are
CC used to determine the identity of the nucleotide at a biallelic marker in
CC a sample DNA sequence. The nucleotide sequences may be labelled and used
CC for genotyping by determining the identity of a nucleotide at a Region D-
CC related biallelic marker in a biological sample from single or multiple
CC subjects. By determining the frequency of a biallelic marker in a
CC population an association between a genotype and a trait, a haplotype and
CC a trait and a phenotype and a trait can be detected. The sequences can be
CC used to determine a predisposition to or early onset of schizophrenia or
CC bipolar disorder or a beneficial response to or side effects related to
CC treatment against schizophrenia or bipolar disorder

SQ Sequence 3001 BP; 952 A; 600 C; 485 G; 963 T; 0 U; 1 Other;

Query Match 79.1%; Score 18.2; DB 3; Length 3001;
Best Local Similarity 87.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCTTTTCATCCCACTACCACTGA 23
DB 1061 CCTGTTCTTCACTACCACTAA 1039

RESULT 2
ACN43926/c
ID ACN43926 standard; DNA; 53242 BP.

XX AC ACN43926;

DT 18-NOV-2004 (first entry)

DE Human genomic sequence HCG1782215.

XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.

OS Homo sapiens.

XX WO2003073826-A2.

PD 12-SEP-2003.

XX 28-FEB-2003; 2003WO-US006235.

XX 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

PA Morris DW;

PI WPI; 2003-328604/31.

DR Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
comprises a nucleotide sequence.

PS Claim 1; SEQ ID NO 118; Opp; English.

XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of

CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published

SQ Sequence 53242 BP; 12636 A; 10850 C; 11247 G; 15524 T; 0 U; 2995 Other;

Query Match 79.1%; Score 18.2; DB 11; Length 53242;
Best Local Similarity 87.0%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCTTTTCATCCCACTACCACTGA 23
DB 10476 CCTTTTCATCCCACTACCACTGA 10454

RESULT 3
ADO79173_0
WP Sequence split into 5 fragments LOCUS ADO79173 Accession ADO79173

WP Fragment Name Begin End
WP ADO79173_0 1 110000
WP ADO79173_1 100001 210000
WP ADO79173_2 200001 310000
WP ADO79173_3 300001 410000
WP ADO79173_4 400001 447894

ID ADO79173 standard; DNA; 447894 BP.

XX AC ADO79173;

XX 26-AUG-2004 (first entry)

DE KLF12 gene and surrounding region, SEQ ID 1.

XX Cytostatic; Gene Therapy; human; breast cancer;

XX zinc finger transcriptional repressor; Kruppel-like factor 12; KLF12;

XX transcription factor; AP-2 alpha; AP-2a; chromosome 13; gene; ds;

XX single nucleotide polymorphism; SNP.

XX Homo sapiens.

XX Location/Qualifiers

FT Key 1
FT variation /tag= a
FT /standard_name= "Biallelic marker"
FT 1437 /tag= b
FT /standard_name= "Biallelic marker"
FT 1604 /tag= c
FT /standard_name= "Biallelic marker"
FT 2251 /tag= d
FT /standard_name= "Biallelic marker"
FT 2906 /tag= e
FT /standard_name= "Biallelic marker"
FT 2953 /tag= f
FT /standard_name= "Biallelic marker"
FT 3131 /tag= g
FT /standard_name= "Biallelic marker"
FT 3133 /tag= h
FT /standard_name= "Biallelic marker"
FT 3238 /tag= i

FT /standard_name= "Biallelic marker"
FT 3594
FT /tag= j
FT /standard_name= "Biallelic marker"
FT 4312
FT /tag= k
FT /standard_name= "Biallelic marker"
FT 4854
FT /tag= l
FT /standard_name= "Biallelic marker"
FT 5100
FT /tag= m
FT /standard_name= "Biallelic marker"
FT 5351
FT /tag= n
FT /standard_name= "Biallelic marker"
FT 5355
FT /tag= o
FT /standard_name= "Biallelic marker"
FT 50479
FT /tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a G/A SNP and the cancer
FT associated allele is G"
FT 50567
FT /tag= q
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a G/A SNP and the cancer
FT associated allele is A"
FT 51865
FT /tag= r
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a G/A SNP and the cancer
FT associated allele is G"
FT 53349
FT /tag= s
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a C/G SNP and the cancer
FT associated allele is G"
FT 59464
FT /tag= t
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a A/T SNP and the cancer
FT associated allele is A"
FT 68551
FT /tag= u
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a C/T SNP and the cancer
FT associated allele is C"
FT 70419
FT /tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a C/G SNP and the cancer
FT associated allele is G"
FT 73014
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a T/G SNP and the cancer
FT associated allele is T"
FT 75242
FT /tag= x
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a G/T SNP and the cancer
FT associated allele is T"
FT 75505
FT /tag= y
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a G/A SNP and the cancer
FT associated allele is A"
FT 78067
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a G/A SNP and the cancer

FT associated allele is G"
FT 78265
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a T/G SNP and the cancer
FT associated allele is G"
FT 78290
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a C/T SNP and the cancer
FT associated allele is C"
FT 78956
FT /tag= ac
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a A/C SNP and the cancer
FT associated allele is A and is also described as a
FT biallelic G/T marker"
FT 79196
FT /tag= ad
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a G/A SNP and the cancer
FT associated allele is G"
FT 81886
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a G/A SNP and the cancer
FT associated allele is A"
FT 82399
FT /tag= af
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a T/G SNP and the cancer
FT associated allele is G"
FT 83380
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a G/A SNP and the cancer
FT associated allele is A"
FT 86645
FT /tag= ah
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a T/C SNP and the cancer
FT associated allele is C"
FT 86704
FT /tag= ai
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a T/C SNP and the cancer
FT associated allele is T"
FT 88134
FT /tag= aj
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a C/G SNP and the cancer
FT associated allele is C"
FT 96416
FT /tag= ak
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a C/A SNP and the cancer
FT associated allele is A"
FT 96429
FT /tag= al
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a A/T SNP and the cancer
FT associated allele is T"
FT 96535
FT /tag= am
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a C/T SNP and the cancer
FT associated allele is C"
FT 110920
FT /tag= an
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a A/C SNP and the cancer
FT associated allele is C"
FT 114117
FT /tag= ao
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "This SNP is described as a A/C SNP and the cancer
FT associated allele is C"

FT /*tag= ao
FT /standard name= "Single nucleotide polymorphism"
FT /notes="This SNP is described as a T/C SNP and the cancer
FT associated allele is T"
FT 117297

Query Match 77.4%; Score 17.8; DB 12; Length 110000;
Best Local Similarity 90.5%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTTTTCATCCAACTACCACTGA 23
 ||||| ||||| |||||
Db 52745 TTTTCATACAACTTCCACTGA 52765

RESULT 4
ACN44848/c
ID ACN44848 standard; DNA; 50287 BP.

XX
AC ACN44848;

XX 18-NOV-2004 (first entry)

XX Mouse genomic sequence mCG4799.

XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.

XX Mus musculus.

XX WO2003073826-A2.

XX 12-SEP-2003.

XX 28-FEB-2003; 2003WO-US006235.

XX 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-328604/31.

XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.

XX Claim 1; SEQ ID NO 1501; Opp; English.

XX The present invention relates to novel DNA and protein sequences which
XX are associated with carcinomas. The sequences are useful for: (i) for
XX screening drug candidates; (ii) for screening of bioactive agent capable
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX a bioactive agent capable of modulating the activity of CAP; (iv) for
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX carcinoma including lymphoma. The present sequence is one such CA coding
XX sequence. Note: This patent is an equivalent to basic patent
XX US2002182586A1, for which no sequence data was published

XX Sequence 50287 BP; 13307 A; 9608 C; 9890 G; 15956 T; 0 U; 1526 Other;

Query Match 75.7%; Score 17.4; DB 11; Length 50287;
Best Local Similarity 94.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTTTTCATCCCAACTACCA 19
 ||||| ||||| |||||
Db 48726 CCTTTTCTTCCCAACTACCA 48708

RESULT 5
ABD33502
ID ABD33502 standard; DNA; 81684 BP.

XX
AC ABD33502;

XX 18-NOV-2004 (first entry)

XX Murine cancer-associated (CA) gene MD07-098.

XX Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX ds; cancer; cytostatic.

XX Mus musculus.

XX WO2004058146-A2.

XX 15-JUL-2004.

XX 15-DEC-2003; 2003WO-US040081.

XX 17-DEC-2002; 2002US-00322281.

XX (SAGR-) SAGRES DISCOVERY INC.

XX Morris DW, Malandro MS;

XX WPI; 2004-499109/47.

XX Novel human cancer associated protein encoded within open reading frame
XX of cancer associated gene, useful as targets for diagnosing cancer.

XX Disclosure; SEQ ID NO 673; 182pp; English.

XX The invention relates to cancer-associated proteins (CAP) and the cancer-
XX associated (CA) nucleic acids encoding them. The invention also relates
XX to a method for treating cancers involving administering to a patient an
XX inhibitor of CAP, and a method of screening for anticancer activity in a
XX potential drug involving providing a cell that expresses a CA gene,
XX contacting a tissue sample derived from a cancer cell with an anticancer
XX drug candidate and monitoring the effect of the anticancer drug candidate
XX on expression of the CA gene. The CAP proteins are useful for detecting
XX cancer associated with expression of a CAP protein in a test cell sample
XX and for screening for a bioactive agent capable of modulating the
XX activity of a CAP protein. The CA nucleic acids are useful for diagnosing
XX cancer, involving determining the expression of a CA nucleic acid in a
XX tissue. This sequence represents a murine CA gene of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 81684 BP; 21861 A; 18581 C; 19236 G; 21459 T; 0 U; 547 Other;

Query Match 75.7%; Score 17.4; DB 13; Length 81684;
Best Local Similarity 94.7%; Pred. No. 6.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTTTTCATCCCAACTACCA 19
 ||||| ||||| |||||
Db 72798 CCTTTTCTTCCCAACTACCA 72816

RESULT 6
AAC28339/c

ID AAC28339 standard; cDNA; 377 BP.

XX
AC AAC28339;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 32414.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX

OS Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX

XX Claim 1; SEQ ID NO 32414; 71pp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors

XX Sequence 377 BP; 83 A; 57 C; 73 G; 159 T; 0 U; 5 Other;
 SQ
 Query Match 74.8%; Score 17.2; DB 3; Length 377;
 Best Local Similarity 86.4%; Pred. No. 4.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTTTTCATCCAACTACCACTGA 23
 Db 288 CATTTCATCCAACTACCACTGA 267

RESULT 7
 ABX38985/c
 ID ABX38985 standard; cDNA; 439 BP.

XX AC ABX38985;

XX 20-FEB-2003 (first entry)

XX Bovine EST associated with lactation/muscle/fat deposition #4150.

XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.

XX Bos Taurus.

XX US2002137139-A1.

XX 26-SEP-2002.

XX 24-SEP-2001; 2001US-00960352.

PR 12-JAN-1999; 99US-0115707P.
 PR 11-JAN-2000; 2000US-00480902.

XX (BYAT/) BYATT J C.

XX (MATH/) MATHIALAGAN N.

XX (TRON/) TAO N.

XX (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI; 2003-110599/10.

XX New nucleic acid associated with lactation, and muscle and fat

PT deposition, useful for genome mapping, gene identification and analysis,

PT cattle breeding, or for genetically improving cattle.

XX

XX Claim 2; SEQ ID NO 4150; 245pp; English.

XX The invention relates to a purified nucleic acid molecule associated with

CC lactation or muscle and fat deposition (designated LMPD), derived from

CC cattle, and the LMPD nucleic acid can specifically hybridise to a second

CC nucleic acid molecule comprising any of 15112 nucleotide sequences,

CC appearing as ABX34836-ABX49947, or complements of them. Also included are

CC ; (1) a transformed cell having a nucleic acid comprising an LMPD nucleic

CC acid linked to a promoter and a 3' non-translated sequence that

CC functions in the cell to cause termination of transcription and addition

CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and

CC (2) determining a level or pattern of a molecule in a bovine cell or

CC tissue comprising a level or pattern of a marker nucleic acid (comprising any

CC of the 15112 nucleic acid sequences or its complement or fragment) with a

CC complementary nucleic acid molecule obtained from the bovine cell or

CC tissue, where hybridisation between the marker nucleic acid and the

CC complementary nucleic acid permits the detection of the molecule; and (b)

CC detecting the level or pattern of the complementary nucleic acid, where

CC the detection of the complementary nucleic acid is predictive of the

CC level or pattern of the molecule. The LMPD nucleic acid is used for

CC determining a level or pattern of a molecule in a bovine cell or tissue.

CC It is useful for genome mapping, gene identification and analysis, cattle

CC breeding, preparation of constructs for use in cattle gene expression, or

CC for genetically improving cattle. The present sequence is one of the

CC 15112 bovine LMPD EST (expressed sequence tag) nucleic acids. Note: The

CC present sequence was not shown in the specification but was obtained in

CC electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139

XX Sequence 439 BP; 152 A; 58 C; 80 G; 149 T; 0 U; 0 Other;

SQ

Query Match 74.8%; Score 17.2; DB 8; Length 439;
 Best Local Similarity 86.4%; Pred. No. 4.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTTTTCATCCAACTACCACTGA 23
 Db 251 CTTTTCATGAACTACACTGA 230

RESULT 8
 ADS59130/c
 ID ADS59130 standard; cDNA; 447 BP.

XX AC ADS59130;

XX 02-DEC-2004 (first entry)

XX Bacterial polynucleotide #11117.

XX Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.

PA (RANK/) RANK D R.
 XX (HANKZ/) HANKZEL D K.
 PI Penn SG, Rank DR, Hanzel DR;
 XX WPI; 2004-119264/12.
 DR
 XX New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 XX
 PS Claim 15; SEQ ID NO 9303; 80pp; English.
 XX
 CC The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 688 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridises under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, and a method of providing
 CC human gene expression data by subcription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterising
 CC alternative splicing events, in detecting and characterising gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704
 XX
 SQ Sequence 588 BP; 253 A; 120 C; 91 G; 124 T; 0 U; 0 Other;
 Query Match 74.8%; Score 17.2; DB 12; Length 588;
 Best Local Similarity 86.4%; Pred. No. 5e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 CTTTTCATCCCACTACACTGA 23
 DB 245 CATTTCATCCCACTACACTG 266
 RESULT 11
 ABQ99614/c
 ID ABQ99614 standard; cDNA; 661 BP.
 XX
 AC ABQ99614;
 XX
 XX 12-NOV-2002 (first entry)
 DT
 XX Human membrane spanning 4-domain family, subfamily A 6E cDNA.
 DE
 XX Human; membrane spanning 4-domain A; cytostatic; antiallergic; MS4A;
 KW gene therapy; atopic disorder; non-Hodgkin's lymphoma;
 XX

KW Hodgkin's lymphoma; allergenic disease; MS4A6E; gene; ss;
 KW chromosome 11q12-13.1.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 15..4586
 FT /*tag= a
 FT /product= "MS4A6E"
 XX
 PN WO200262946-A2.
 XX
 XX 15-AUG-2002.
 PD
 XX 10-DEC-2001; 2001WO-US048437.
 PF
 XX 08-DEC-2000; 2000US-0254362P.
 PR
 XX 20-FEB-2001; 2001US-0270057P.
 PR
 XX (UYDU-) UNIV DUKE.
 PA
 XX Tedder TF, Liang YH;
 PI
 XX WPI; 2002-857530/70.
 DR P-PSDB; ABP65028.
 DR
 XX New membrane spanning 4-domain A (MS4A) genes and polypeptides, useful
 PT for generating animal models of atopic disorders, for drug screening, or
 PT for treating (non-) Hodgkin's lymphoma, or allergenic or atopic disorders
 PT in e.g. humans.
 PT
 XX Claim 2; Page 125-126; 450pp; English.
 PS
 XX The invention relates to novel membrane spanning 4-domain A (MS4A)
 CC nucleic acid and polypeptide molecules, comprising human and mouse MS4A.
 CC The polypeptides of the invention have cytostatic and antiallergic
 CC activity. The polynucleotides may have a use in gene therapy. The MS4A
 CC nucleic acids and polypeptides are useful for generating animal (e.g.
 CC mouse) models of atopic disorders, or for drug discovery screens. These
 CC are also useful for treating (non-)Hodgkin's lymphoma, allergenic
 CC diseases, atopic disorders or other MS4A-related conditions. The present
 CC sequence encodes human membrane spanning 4-domain A 6E (MS4A6E)
 CC
 SQ Sequence 661 BP; 195 A; 147 C; 132 G; 187 T; 0 U; 0 Other;
 Query Match 74.8%; Score 17.2; DB 6; Length 661;
 Best Local Similarity 86.4%; Pred. No. 5e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 CTTTTCATCCCACTACACTG 22
 DB 285 CTTTTCGTCCACTTACACTG 264
 RESULT 12
 AAC93484
 ID AAC93484 standard; cDNA; 728 BP.
 XX
 AC AAC93484;
 XX
 XX 16-FEB-2001 (first entry)
 DT
 XX Human secreted protein gene 6 SEQ ID NO:16.
 DE
 XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiac; vasotropic; cerebroprotective;
 KW notropic; neuroprotective; antibacterial; virucide; fungicide;
 KW ophthalmological; vulnary; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorders; cancer; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
 KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.
 XX
 OS Homo sapiens.

XX WO200061626-A1.
 PN 19-OCT-2000.
 XX 06-APR-2000; 2000WO-US009066.
 XX 09-APR-1999; 99US-0128698P.
 PR 20-JAN-2000; 2000US-0176926P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX Rosen CA, Ruben SM, Komatsoulis G;
 PI WPI; 2000-619227/59.
 XX DR P-PSDB; AAB51832.
 XX New nucleic acid molecules encoding 49 human secreted proteins for
 PT diagnosing, preventing or ameliorating medical conditions and used for
 PT food additives or preservatives.
 XX Claim 1; Page 435-436; 516pp; English.
 PS Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
 CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
 CC AAB51927 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences with which they share homology. The genes and
 CC proteins have activities dependent on the tissues and cells in which they
 CC are expressed. Examples of their activities include immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiac;
 CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
 CC virucide; fungicide; opthalmological; and vulnary. The secreted
 CC proteins, polynucleotides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angioneuroma, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional
 CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
 CC used in the isolation and characterisation of the proteins and
 CC polynucleotides of the invention
 XX SQ Sequence 728 BP; 178 A; 169 C; 183 G; 198 T; 0 U; 0 Other;
 Query Match 74.8%; Score 17.2; DB 3; Length 728;
 Best Local Similarity 86.4%; Pred. No. 5.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 CCTTTTCATCCCACTTACCCTG 22
 ||||| ||||| ||||| ||||| |||||
 Db 178 CCTGTGTCAGCAACTTCCACTG 199
 RESULT 13
 ABZ73565
 ID ABZ73565 standard; cDNA; 728 BP.
 XX ABZ73565;
 XX 12-MAY-2003 (first entry)
 XX Secreted protein-encoding gene 285 cDNA clone HRDFK37, SEQ ID NO:295.
 DE Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 XX

KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianaemic; vulnary; gene; ss.
 XX Homo sapiens.
 XX WO200277013-A2.
 PN 03-OCT-2002.
 PD 26-MAR-2002; 2002WO-US009370.
 PF 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM;
 PI WPI; 2003-040578/03.
 XX DR P-PSDB; ABR01231.
 XX New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
 XX Claim 21; Page 1307; 2474pp; English.
 PS ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein-encoding cDNA clone of the invention
 XX SQ Sequence 728 BP; 178 A; 169 C; 183 G; 198 T; 0 U; 0 Other;
 Query Match 74.8%; Score 17.2; DB 8; Length 728;
 Best Local Similarity 86.4%; Pred. No. 5.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 CCTTTTCATCCCACTTACCCTG 22
 ||||| ||||| ||||| ||||| |||||
 Db 178 CCTGTGTCAGCAACTTCCACTG 199
 RESULT 14
 ADA98080
 ID ADA98080 standard; cDNA; 728 BP.
 XX ADA98080;
 XX 20-NOV-2003 (first entry)
 XX Human secreted protein cDNA sequence #174.
 DE

XX human; secreted protein; cardiovascular disorder; arrhythmia;
 KW atherosclerosis; stroke; endocarditis; congestive heart failure;
 KW rheumatic heart disease; cardiomyopathy; haemorrhoids; varicose veins;
 KW migraine; thrombosis; neural disorder; immune system disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; renal disorder; proliferative disorder; cancer; gene;
 XX ss.
 XX Homo sapiens.
 OS
 XX
 XX WQ2003004623-A2.
 PN
 XX
 XX 16-JAN-2003.
 PD
 XX
 XX 26-MAR-2002; 2002WO-US009922.
 PF
 XX
 XX 27-MAR-2001; 2001US-0278650P.
 PR
 XX 12-SEP-2001; 2001US-00950082.
 PR
 XX 12-SEP-2001; 2001US-00950083.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Rosen CA, Ruben SM;
 PI
 XX WPI; 2003-247946/24.
 DR
 XX
 XX New human secreted polypeptide and nucleic acid molecules, useful for
 PT diagnosing, preventing, prognosticating or treating cardiovascular
 PT disorders (e.g. arrhythmia, atherosclerosis, cardiomyopathy, or
 PT thrombosis).
 XX
 XX Claim 1; SEQ ID NO 184; 1572pp; English.
 PS
 XX The invention comprises the amino acid and coding sequence of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful in the treatment of cardiovascular disorders, such as: arrhythmia,
 CC atherosclerosis, stroke, endocarditis, congestive heart failure,
 CC rheumatic heart disease, cardiomyopathy, haemorrhoids, varicose veins,
 CC migraine, or thrombosis. The DNA and protein sequences may also be used
 CC for treating or preventing: neural disorders, immune system disorders,
 CC muscular disorders, reproductive disorders, gastrointestinal disorders,
 CC pulmonary disorders, renal disorders, proliferative disorders and/or
 CC cancerous diseases. The present cDNA sequence encodes a human secreted
 CC protein of the invention. NOTE: The present sequence is shown on the WIPO
 CC website.
 XX
 XX Sequence 728 BP; 178 A; 169 C; 183 G; 198 T; 0 U; 0 Other;
 SQ
 Query Match 74.8%; Score 17.2; DB 8; Length 728;
 Best Local Similarity 86.4%; Pred. No. 5.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 CCTTTTCATCCCACTACCACTG 22
 |||||
 Db 178 CCTTGTGAGCCCACTTCCACTG 199
 |||||
 RESULT 15
 ABZ67162
 ID ABZ67162 standard; cDNA; 728 BP.
 XX
 XX AC ABZ67162;
 XX
 XX 26-MAR-2003 (first entry)
 DT
 XX
 XX Human secreted protein encoding cDNA SEQ ID NO 282.
 DE
 XX
 XX Human; secreted protein; neurotropic; neuroprotective; cytostatic;
 KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW vulvarity; antibacterial; antiparkinsonian; antisickling; antianaemic;
 KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
 KW antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;

KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; nephrotropic;
 XX gene therapy; gene; ds.
 OS
 XX Homo sapiens.
 XX WQ200277186-A2.
 PN
 XX
 XX 03-OCT-2002.
 PD
 XX
 XX 26-MAR-2002; 2002WO-US009188.
 PF
 XX
 XX 27-MAR-2001; 2001US-0278650P.
 PR
 XX 12-SEP-2001; 2001US-00950082.
 PR
 XX 12-SEP-2001; 2001US-00950083.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Rosen CA, Ruben SM;
 PI
 XX WPI; 2003-040583/03.
 DR
 XX P-PSDB; ABP99741.
 DR
 XX
 XX New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
 PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
 PT West Nile fever.
 XX
 XX Claim 7; Page 1316; 2423pp; English.
 PS
 XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections.
 XX
 XX Sequence 728 BP; 178 A; 169 C; 183 G; 198 T; 0 U; 0 Other;
 SQ
 Query Match 74.8%; Score 17.2; DB 10; Length 728;
 Best Local Similarity 86.4%; Pred. No. 5.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 CCTTTTCATCCCACTACCACTG 22
 |||||
 Db 178 CCTTGTGAGCCCACTTCCACTG 199
 |||||

Search completed: August 13, 2005, 04:14:35
 Job time : 191.245 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 03:33:32 ; Search time 56.5213 Seconds
(without alignments)
665.844 Million cell updates/sec

Title: US-10-673-854-2

Perfect score: 23

Sequence: 1 cctttcatccactaccactga 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| C 1 | 18.2 | 79.1 | 3001 | 4 | US-09-539-333D-191 |
| C 2 | 17.8 | 77.4 | 103447 | 4 | Sequence 191, App |
| C 3 | 17.2 | 74.8 | 377 | 4 | Sequence 16320, A |
| C 4 | 17.2 | 74.8 | 513 | 4 | Sequence 32414, A |
| C 5 | 17.2 | 74.8 | 562 | 4 | Sequence 1795, Ap |
| C 6 | 17.2 | 74.8 | 601 | 4 | Sequence 17125, A |
| C 7 | 17.2 | 74.8 | 601 | 4 | Sequence 96155, A |
| C 8 | 17.2 | 74.8 | 2061 | 4 | Sequence 96156, A |
| C 9 | 17.2 | 74.8 | 4770 | 4 | Sequence 366, App |
| C 10 | 17.2 | 74.8 | 36755 | 4 | Sequence 826, App |
| C 11 | 17.2 | 74.8 | 103750 | 4 | Sequence 16994, A |
| C 12 | 17.2 | 74.8 | 162450 | 3 | Sequence 13319, A |
| C 13 | 17.2 | 74.8 | 387902 | 4 | Sequence 1, Appli |
| C 14 | 17.2 | 74.8 | 421883 | 4 | Sequence 14543, A |
| C 15 | 16.8 | 73.0 | 942 | 4 | Sequence 12557, A |
| C 16 | 16.8 | 73.0 | 15478 | 4 | Sequence 3203, Ap |
| C 17 | 16.8 | 73.0 | 74644 | 4 | Sequence 15288, A |
| C 18 | 16.8 | 73.0 | 784019 | 4 | Sequence 17556, A |
| C 19 | 16.8 | 73.0 | 828152 | 4 | Sequence 14033, A |
| C 20 | 16.6 | 72.2 | 601 | 4 | Sequence 12777, A |
| C 21 | 16.6 | 72.2 | 601 | 4 | Sequence 199769, A |
| C 22 | 16.6 | 72.2 | 601 | 4 | Sequence 199770, A |
| C 23 | 16.6 | 72.2 | 325791 | 4 | Sequence 199771, A |
| C 24 | 16.6 | 72.2 | 373182 | 4 | Sequence 1, Appli |
| C 25 | 16.6 | 72.2 | 373694 | 4 | Sequence 17371, A |
| C 26 | 16.4 | 71.3 | 601 | 4 | Sequence 12062, A |
| C 27 | 16.4 | 71.3 | 47115 | 4 | Sequence 169822, A |
| C 28 | 16.4 | 71.3 | 47122 | 4 | Sequence 12278, A |

ALIGNMENTS

RESULT 1

US-09-539-333D-191/c

; Sequence 191, Application US/09539333D

; Patent No. 6476208

; GENERAL INFORMATION:

; APPLICANT: Cohen, Daniel

; APPLICANT: Blumenfeld, Marta

; APPLICANT: Chumakov, Ilya

; APPLICANT: Bougueleret, Lydie

; APPLICANT: Bihaïn, Bernard

; APPLICANT: Essioux, Laurent

; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS

; FILE REFERENCE: GENSET.047AUS

; CURRENT APPLICATION NUMBER: US/09/539,333D

; CURRENT FILING DATE: 2000-03-30

; PRIOR APPLICATION NUMBER: US 60/126,903

; PRIOR FILING DATE: 1999-03-30

; PRIOR APPLICATION NUMBER: US 60/131,971

; PRIOR FILING DATE: 1999-04-30

; PRIOR APPLICATION NUMBER: US 60/132,065

; PRIOR FILING DATE: 1999-04-30

; PRIOR APPLICATION NUMBER: US 60/143,928

; PRIOR FILING DATE: 1999-07-14

; PRIOR APPLICATION NUMBER: US 60/145,915

; PRIOR FILING DATE: 1999-07-27

; PRIOR APPLICATION NUMBER: US 60/146,453

; PRIOR FILING DATE: 1999-07-29

; PRIOR APPLICATION NUMBER: US 60/146,452

; PRIOR FILING DATE: 1999-07-29

; PRIOR APPLICATION NUMBER: US 60/162,288

; PRIOR FILING DATE: 1999-10-28

; PRIOR APPLICATION NUMBER: US 09/416,384

; PRIOR FILING DATE: 1999-10-12

; NUMBER OF SEQ ID NOS: 231

; SOFTWARE: Patent.pm

; SEQ ID NO 191

; LENGTH: 3001

; TYPE: DNA

; ORGANISM: Homo Sapiens

; FEATURE:

; NAME/KEY: allele

; LOCATION: 1501

; OTHER INFORMATION: 99-26191-58 : polymorphic base G or A

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1502..1520

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

; OTHER INFORMATION: 99-26191-58.misl, complement

; FEATURE:

; NAME/KEY: misc binding

; LOCATION: 1481..1500

```
; OTHER INFORMATION: 99-26191-58.mis2,
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1539..1558
; OTHER INFORMATION: upstream amplification primer, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1095..1115
; OTHER INFORMATION: downstream amplification primer
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1489..1513
; OTHER INFORMATION: 99-26191-58 probe
US-09-539-333D-191

Query Match          79.1%; Score 18.2; DB 4; Length 3001;
Best Local Similarity 87.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCTTTTCATCCCACTACCACTGA 23
   ||||| ||||| ||||| ||||| |||||
Db 1061 CCGTTCATTCACCACTACCACTAA 1039

RESULT 2
US-09-949-016-16320/c
; Sequence 16320, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16320
; LENGTH: 103447
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(103447)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16320

Query Match          77.4%; Score 17.8; DB 4; Length 103447;
Best Local Similarity 90.5%; Pred. No. 11e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTTTCATCCCACTACCACTGA 23
   ||||| ||||| ||||| ||||| |||||
Db 7931 TTTCCATCAAACTACCACTGA 7911

RESULT 3
US-09-513-999C-32414/c
; Sequence 32414, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
```

```
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 32414
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 144_feature
; OTHER INFORMATION: r-a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 235
; OTHER INFORMATION: r-a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 236
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 261
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 335
; OTHER INFORMATION: d=a or g or t
US-09-513-999C-32414

Query Match          74.8%; Score 17.2; DB 4; Length 377;
Best Local Similarity 86.4%; Pred. No. 95;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTTTTCATCCCACTACCACTGA 23
   | ||||| ||||| ||||| |||||
Db 288 CATTTCATCCCAACCAACACAGA 267

RESULT 4
US-09-621-976-1795/c
; Sequence 1795, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1795
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5..409
; NAME/KEY: sig_peptide
; LOCATION: 5..181
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 4.5
; OTHER INFORMATION: seq LSALSALVGFILL/SV
US-09-621-976-1795

Query Match          74.8%; Score 17.2; DB 4; Length 513;
Best Local Similarity 86.4%; Pred. No. 99;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Qy 1 CCTTTTCATCCAACTACCACTG 22
|||||
Db 239 CCTTTTCGTCCAACTTACACTG 218

RESULT 5
US-09-621-976-17125/c
; Sequence 17125, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S. Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 17125
; LENGTH: 562
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-17125

Query Match 74.8%; Score 17.2; DB 4; Length 562;
Best Local Similarity 86.4%; Pred. No. 1e+02; 3; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CCTTTTCATCCAACTACCACTG 22
|||||
Db 331 CCTTTTCGTCCAACTTACACTG 310

RESULT 6
US-09-949-016-96155
; Sequence 96155, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 96155
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-96155

Query Match 74.8%; Score 17.2; DB 4; Length 601;
Best Local Similarity 86.4%; Pred. No. 1e+02; 3; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCTTTTCATCCAACTACCACTG 23
|||||
Db 368 CATTTTCACCAACACACACAGA 389

RESULT 7
US-09-949-016-96156
; Sequence 96156, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 96156
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-96156

Query Match 74.8%; Score 17.2; DB 4; Length 601;
Best Local Similarity 86.4%; Pred. No. 1e+02; 3; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCTTTTCATCCAACTACCACTG 23
|||||
Db 438 CATTTTCACCAACACACACAGA 459

RESULT 8
US-09-774-528-366/c
; Sequence 366, Application US/09774528
; Patent No. 6743619
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Yang, Yonghong
; APPLICANT: Xue, Aigong J.
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6743619el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/774,528
; CURRENT FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 366
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (97)..(1206)
US-09-774-528-366

Query Match 74.8%; Score 17.2; DB 4; Length 2061;
Best Local Similarity 86.4%; Pred. No. 1.2e+02; 3; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CCTTTTCATCCAACTACCACTG 22
|||||
Db 53 CCTTTTCGTCCAACTTACACTG 32

```
RESULT 9
US-09-543-681A-826/c
; Sequence 826, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 826
; LENGTH: 4770
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-826

Query Match          74.8%; Score 17.2; DB 4; Length 4770;
Best Local Similarity 86.4%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CTTTTCATCCAACTACCACTGA 23
Db      1988 CTTTTCATCCAACTACCAATGA 1967

RESULT 10
US-09-949-016-16994
; Sequence 16994, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16994
; LENGTH: 36755
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16994

Query Match          74.8%; Score 17.2; DB 4; Length 36755;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CCTTTTCATCCAACTACCACTG 22
Db      26383 CCTTTCACCAACACCACTG 26404

RESULT 11
US-09-949-016-13319/c
; Sequence 13319, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13319
; LENGTH: 103750
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13319

Query Match          74.8%; Score 17.2; DB 4; Length 103750;
Best Local Similarity 86.4%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CTTTTCATCCAACTACCACTGA 23
Db      26131 CTTTTCATCCAAAGGACCAATGA 26110

RESULT 12
US-09-345-882-1/c
; Sequence 1, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; FILE REFERENCE: GENSET.031A
; CURRENT APPLICATION NUMBER: US/09/345,882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 162450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72794
; OTHER INFORMATION: 5-124-273 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88073
; OTHER INFORMATION: 5-127-261 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90842
; OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93714
; OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97122
; OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97152
; OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99098
; OTHER INFORMATION: 5-130-257 : polymorphic base A or G
```



```
,
,
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99117
, OTHER INFORMATION: 5-130-276 : polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 103806
, OTHER INFORMATION: 5-131-395 : polymorphic base A or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 106940
, OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108106
, OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108149
, OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108308
, OTHER INFORMATION: 5-135-357 : polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108471
, OTHER INFORMATION: 5-136-174 : polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 134134
, OTHER INFORMATION: 5-140-120 : polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 134362
, OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
, FEATURE:
, NAME/KEY: allele
, LOCATION: 134374
, OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
, FEATURE:
, NAME/KEY: allele
, LOCATION: 146328
, OTHER INFORMATION: 5-143-84 : polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 146345
, OTHER INFORMATION: 5-143-101 : polymorphic base A or C
, FEATURE:
, NAME/KEY: allele
, LOCATION: 150329
, OTHER INFORMATION: 5-145-24 : polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 160031
, OTHER INFORMATION: 5-148-352 : polymorphic base G or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 72771..72817
, OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
, FEATURE:
, NAME/KEY: allele
, LOCATION: 72771..72817
, OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
, FEATURE:
, NAME/KEY: allele
, LOCATION: 88050..88096
, OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
, FEATURE:
, NAME/KEY: allele
, LOCATION: 88050..88096
, OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
, FEATURE:
, NAME/KEY: allele
, LOCATION: 90819..90865
, OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
, FEATURE:
, NAME/KEY: allele
, LOCATION: 90819..90865
, OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
, FEATURE:
, NAME/KEY: allele
, LOCATION: 93690..93736
, OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
, FEATURE:
, NAME/KEY: allele
, LOCATION: 93690..93736
, OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97099..97145
, OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97099..97145
, OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97130..97177
, OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
, FEATURE:
, NAME/KEY: allele
, LOCATION: 97130..97177
, OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99075..99121
, OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99075..99121
, OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99094..99140
, OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
, FEATURE:
, NAME/KEY: allele
, LOCATION: 99094..99140
, OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
, FEATURE:
, NAME/KEY: allele
, LOCATION: 103783..103828
, OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
, FEATURE:
, NAME/KEY: allele
, LOCATION: 103783..103828
, OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
, FEATURE:
, NAME/KEY: allele
, LOCATION: 106918..106966
, OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
, FEATURE:
, NAME/KEY: allele
, LOCATION: 106918..106966
, OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108084..108130
, OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
, FEATURE:
, NAME/KEY: allele
, LOCATION: 108084..108130
, OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
, FEATURE:
, NAME/KEY: allele
```

```
; LOCATION: 108127..108177
; OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108127..108177
; OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
; FEATURE:
```

```
Query Match          74.8%; Score 17.2; DB 3; Length 162450;
Best Local Similarity 86.4%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 CTTTTCATCCAACTACCACTGA 23
Db      76273 CTATTCATCCAACTACCACTTA 76252
```

```
RESULT 13
US-09-949-016-14543/c
; Sequence 14543, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14543
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)_(387902)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14543
```

```
Query Match          74.8%; Score 17.2; DB 4; Length 387902;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 CTTTTCATCCAACTACCACTGA 23
Db      187998 CATTTCATCCAACTACCACTGA 187977
```

```
RESULT 14
US-09-949-016-12557/c
; Sequence 12557, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12557
; LENGTH: 421883
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(421883)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12557
```

```
Query Match          74.8%; Score 17.2; DB 4; Length 421883;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 CTTTTCATCCAACTACCACTGA 23
Db      187998 CATTTCATCCAACTACCACTGA 187977
```

```
RESULT 15
US-09-328-352-3203/c
; Sequence 3203, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3203
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3203
```

```
Query Match          73.0%; Score 16.8; DB 4; Length 942;
Best Local Similarity 90.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      3 TTTTCATCCAACTACCACTG 22
Db      22 TTTTCATCGAATTACCACTG 3
```

```
Search completed: August 13, 2005, 06:48:49
Job time : 61.5213 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 03:26:33 ; Search time 1478.12 Seconds
(without alignments)
592.293 Million cell updates/sec

Title: US-10-673-854-2

Perfect score: 23

Sequence: 1 cctttcatcctaactaccatga 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 20 | 87.0 | 1018 | 8 | BH163891 ENTTC11TF |
| 2 | 19.4 | 84.3 | 592 | 1 | AJ631826 AJ631826 |
| 3 | 18.8 | 81.7 | 404 | 2 | BF048570 dc30f02.x |
| 4 | 18.8 | 81.7 | 627 | 8 | AZ334397 IMW0063G14 |
| 5 | 18.8 | 81.7 | 644 | 6 | CB102377 MWV_SQ008 |
| 6 | 18.8 | 81.7 | 860 | 9 | CL072324 CH216-124 |
| 7 | 18.4 | 80.0 | 529 | 8 | AZ926545 476_gis08 |
| 8 | 18.2 | 79.1 | 204 | 1 | AA073416 mm84802.r |
| 9 | 18.2 | 79.1 | 228 | 6 | CD067162 MA1-0049T |
| 10 | 18.2 | 79.1 | 281 | 2 | BB486889 BB486889 |
| 11 | 18.2 | 79.1 | 295 | 1 | AV107656 AV107656 |
| 12 | 18.2 | 79.1 | 394 | 8 | BH170823 SALK_0033 |
| 13 | 18.2 | 79.1 | 452 | 6 | CB101123 kl66g09.y |
| 14 | 18.2 | 79.1 | 581 | 9 | CE706312 tigr-gss- |
| 15 | 18.2 | 79.1 | 1184 | 8 | CC320497 TAM32-27E |
| 16 | 18.2 | 79.1 | 2145 | 3 | AK044653 Mus muscu |
| 17 | 18 | 78.3 | 649 | 9 | CG822989 SOYFG92TH |
| 18 | 17.8 | 77.4 | 279 | 3 | CNS09JH9 |
| 19 | 17.8 | 77.4 | 404 | 8 | AQ020753 CIT-HSP-2 |
| 20 | 17.8 | 77.4 | 410 | 3 | CNS09JBW |
| 21 | 17.8 | 77.4 | 422 | 8 | AZ456776 Single re |
| 22 | 17.8 | 77.4 | 497 | 8 | B63439 |
| 23 | 17.8 | 77.4 | 500 | 8 | AQ586167 RPEC11-4 |
| 24 | 17.8 | 77.4 | 528 | 1 | AL926454 AL926454 |

| | | | | | | |
|----|------|------|------|---|----------|--------------------|
| 25 | 17.8 | 77.4 | 574 | 1 | AV769356 | AV769356 AV769356 |
| 26 | 17.8 | 77.4 | 576 | 4 | BM646088 | BM646088 170006873 |
| 27 | 17.8 | 77.4 | 577 | 7 | CF447387 | CF447387 EST683732 |
| 28 | 17.8 | 77.4 | 672 | 9 | CB319820 | CB319820 Medicago |
| 29 | 17.8 | 77.4 | 722 | 6 | CB818885 | CB818885 EST 1775 |
| 30 | 17.8 | 77.4 | 752 | 9 | AG376278 | AG376278 Mus muscu |
| 31 | 17.8 | 77.4 | 816 | 6 | CB230720 | CB230720 AGENCOURT |
| 32 | 17.8 | 77.4 | 866 | 3 | CNS09P2S | EX068576 Single re |
| 33 | 17.8 | 77.4 | 868 | 9 | EX983278 | EX983278 Forward s |
| 34 | 17.8 | 77.4 | 877 | 3 | CNS09P2R | EX068575 Single re |
| 35 | 17.8 | 77.4 | 884 | 7 | CR534985 | CR534985 CR534985 |
| 36 | 17.8 | 77.4 | 931 | 3 | CNS09P2Y | EX068870 Single re |
| 37 | 17.8 | 77.4 | 947 | 3 | CNS09S15 | EX043209 Single re |
| 38 | 17.8 | 77.4 | 954 | 3 | CNS093GM | EX040562 Single re |
| 39 | 17.8 | 77.4 | 954 | 8 | CC208750 | CC208750 CH261-101 |
| 40 | 17.8 | 77.4 | 985 | 9 | AY412408 | AY412408 Pan trogl |
| 41 | 17.8 | 77.4 | 1022 | 3 | CNS09PAX | EX068869 Single re |
| 42 | 17.8 | 77.4 | 1053 | 3 | CNS09J6C | EX060928 Single re |
| 43 | 17.8 | 77.4 | 1122 | 4 | BM563746 | BM563746 AGENCOURT |
| 44 | 17.4 | 75.7 | 220 | 9 | CG590120 | CG590120 OST243140 |
| 45 | 17.4 | 75.7 | 246 | 1 | AV360610 | AV360610 AV360610 |

ALIGNMENTS

RESULT 1
BH163891 1018 bp DNA linear GSS 24-SBP-2001
LOCUS ENTTC11TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, genomic survey sequence.
ACCESSION BH163891
VERSION BH163891.1 GI:15737329
KEYWORDS GSS.
SOURCE Entamoeba histolytica
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 1018)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HM1:IMSS sheared DNA library (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 490.
Location/Qualifiers
1..1018
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site 1: Bat I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

```

ORIGIN
Query Match      87.0%; Score 20; DB 8; Length 1018;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTTTCATCCCACTACCAC 20
    |||||
Db 777 CCTTTTCATCCCACTACCAC 796

RESULT 2
AJ631826/c
LOCUS      AJ631826 Prunus persica mesocarp S4 climacteric Prunus persica cDNA
DEFINITION clone Pp-S4EST0157, mRNA sequence.
ACCESSION  AJ631826
VERSION     AJ631826.1 GI:46472729
SOURCE      EST.
ORGANISM    Prunus persica (peach)
            Prunus persica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
            1 (bases 1 to 592)
            Zilicotto,F., Begheldo,M., Rasori,A., Bonghi,C., Ramina,A. and
            Tonutti,P.
            Microarray to study molecular and genetic aspects of ripening and
            qualitative traits of peach (Prunus persica L. batesch) fruit
            Unpublished (2004)
JOURNAL
COMMENT     Contact: Tonutti P
            Environmental Agronomy and Crop Science
            University of Padova
            viale dell'Università, 16 Agripolis, Legnaro (PD) /35020, ITALY.

FEATURES
source
1..592
    /organism="Prunus persica"
    /mol_type="mRNA"
    /cultivar="Pantasia"
    /db_xref="taxon:3760"
    /clone="Pp-S4EST0157"
    /tissue_type="mesocarp"
    /dev_stage="S4 climacteric"
    /clone_lib="Prunus persica mesocarp S4 climacteric"

ORIGIN
Query Match      84.3%; Score 19.4; DB 1; Length 592;
Best Local Similarity 95.2%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CTTTTCATCCCACTACCAC 22
    |||||
Db 419 CATTTCATCCCACTACCAC 399

RESULT 3
BF048570
LOCUS      BF048570 404 bp mRNA linear EST 11-OCT-2000
DEFINITION 3', mRNA sequence.
ACCESSION  BF048570
VERSION     BF048570.1 GI:10767073
KEYWORDS   EST.
SOURCE      Xenopus laevis (African clawed frog)
ORGANISM    Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
            1 (bases 1 to 404)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 403.
Location/Qualifiers
1..404
    /organism="Xenopus laevis"
    /mol_type="mRNA"
    /db_xref="taxon:8355"
    /clone="IMAGE:3398619"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NICHD_XGC_Lil"
    /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
    Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
    Average insert size 1.4 kb. Constructed by Life
    Technologies. Note: This is a Xenopus Gene Collection
    (XGC) library."

ORIGIN
Query Match      81.7%; Score 18.8; DB 2; Length 404;
Best Local Similarity 90.9%; Pred. No. 4.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCTTTTCATCCCACTACCAC 22
    |||||
Db 319 CCTTTTCATCCCACTACCAC 340

RESULT 4
AZ334397/c
LOCUS      AZ334397 627 bp DNA linear GSS 29-SEP-2000
DEFINITION clone UUGC1M063G14 R, genomic survey sequence.
ACCESSION  AZ334397
VERSION     AZ334397.1 GI:10401685
KEYWORDS   GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 627)
            Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
JOURNAL
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0063 row: G column: 14
            Seq primer: CACACGAGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 627.
            Location/Qualifiers
            1..627
            /organism="Mus musculus"
            /mol_type="genomic DNA"

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0063G14"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/notes="Vector: FWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydronamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (G14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

ORIGIN

```

Query Match      81.7%; Score 18.8; DB 8; Length 627;
Best Local Similarity 90.9%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 CTTTTCATCCAACCTACCACTGA 23
|||||
Db 70 CTTTTCATCCAACCTACCACTGA 49
|||||

```

RESULT 5

```

LOCUS CB102377 644 bp mRNA linear EST 28-JAN-2003
DEFINITION MWV_SQ008044 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
ACCESSION CB102377
VERSION CB102377.1 GI:27928184
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans

```

REFERENCE

```

AUTHORS Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
TITLE Protein interaction mapping in C. elegans using proteins involved
in vulval development

```

JOURNAL

```

MEDLINE Science 287 (5450), 116-122 (2000)
PUBMED 20082953
COMMENT Contact: Vidal M
Dana Farber Cancer Institute
10615043

```

```

Contact: Vidal M
Dana Farber Cancer Institute
10615043

```

```

1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace dvp10604.x with Bait unknown
POLYA=No.

```

FEATURES

```

source
1. .644
Location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"

```

```

/clone_lib="AD-wrmcDNA"
/notes="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"

```

ORIGIN

```

Query Match      81.7%; Score 18.8; DB 6; Length 644;
Best Local Similarity 90.9%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 CTTTTCATCCAACCTACCACTGA 23
|||||
Db 164 CTTTTCATCCAACCTACCACTGA 185
|||||

```

RESULT 6

```

LOCUS CL072324 860 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-124E11 RM1.1 CH216 Xenopus tropicalis genomic clone
CH216-124E11, genomic survey sequence.
ACCESSION CL072324
VERSION CL072324.1 GI:40528237
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;
Xenopodinae; Xenopus; Silurana.

```

```

1 (bases 1 to 860)
Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM1 TAGCACTCACTATAGGAGA
Class: BAC ends
High quality sequence start: 102
High quality sequence stop: 445.

```

FEATURES

```

source
1..860
Location/Qualifiers
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-124E11"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/notes="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"

```

ORIGIN

```

Query Match      81.7%; Score 18.8; DB 9; Length 860;
Best Local Similarity 90.9%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 CTTTTCATCCAACCTACCACTG 22
|||||
Db 798 CTTTTCATCCAACCTACCACTG 819
|||||

```

RESULT 7

```

LOCUS AZ926545/c 529 bp DNA linear GSS 01-APR-2001
DEFINITION 476.d1808d12.s1 Saccharomyces castellii NRRL Y-12630 Naumovia
castellii genomic clone 476.d1808d12.s1, genomic survey sequence.
ACCESSION AZ926545

```

| | | | |
|-----------------------|---|--------------------------------------|---------------------------------|
| VERSION | AZ926545.1 | GI:13497447 | |
| KEYWORDS | GSS. | | |
| SOURCE | Naumovia castellii | | |
| ORGANISM | Naumovia castellii | | |
| REFERENCE | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Naumovia. | | |
| AUTHORS | 1 (bases 1 to 529) Clifton,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish,W.R., Waterston,R.H. and Johnston,M. | | |
| TITLE | Surveying Saccharomycetes genomes to identify functional elements by comparative DNA sequence analysis | | |
| JOURNAL | Unpublished (2001) | | |
| COMMENT | Contact: Johnston M Department of Genetics Washington University Medical School Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA Tel: 314 362 2735 Fax: 314 362 7855 Email: mj@genetics.wustl.edu Class: random plasmid subclone. | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..529 /organism="Naumovia castellii" /mol_type="genomic DNA" /strain="NRRL Y-12630 (CBS 4309)" /db_xref="taxon:27288" /clone="476.dis08d12.s1" /clone.lib="Saccharomycetes castellii NRRL Y-12630" /note="Random genomic sequence " | | |
| ORIGIN | | | |
| Query Match | 80.0%; | Score 18.4; | DB 8; |
| Best Local Similarity | 95.0%; | Pred. No. 7,7e+02; | |
| Matches | 19; | Conservative 0; | Mismatches 1; Indels 0; Gaps 0; |
| QY | 4 | TTTCATCCCACTACCCTGA 23 | |
| Db | 520 | TTTCATCCCAAGTACCCTGA 501 | |
| RESULT 8 | | | |
| AA073416 | | | |
| LOCUS | mm84a02.r1 | Stratagene mouse embryonic carcinoma | EST 07-FEB-1997 |
| DEFINITION | musculus cDNA clone IMAGE:535082 5', mRNA sequence. | | |
| ACCESSION | AA073416 | | |
| VERSION | AA073416.1 | GI:1595145 | |
| KEYWORDS | EST. | | |
| SOURCE | Mus musculus (house mouse) | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| REFERENCE | 1 (bases 1 to 204) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,H. | | |
| AUTHORS | The WashU-HMI Mouse EST Project Unpublished (1996) Contact: Marra M/Mouse EST Project | | |
| TITLE | WashU-HMI Mouse EST Project | | |
| JOURNAL | Washington University School of MedicineP | | |
| COMMENT | 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MG1:322018 Seq primer: -28mi3 rev1 ET from Amersham High quality sequence stop: 84. | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..204 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:535082" /tissue_type="carcinoma" /dev_stage="embryonic" /lab_host="SOLR (kanamycin resistant)" /clone.lib="Stratagene mouse embryonic carcinomaRA (#937318)" /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dr. P19 cell line treated with retinoic acid. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTT 3" | | |
| ORIGIN | | | |
| Query Match | 79.1%; | Score 18.2; | DB 1; |
| Best Local Similarity | 87.0%; | Pred. No. 8,4e+02; | |
| Matches | 20; | Conservative 0; | Mismatches 3; Indels 0; Gaps 0; |
| QY | 1 | CTTTTCATCCCACTACCCTGA 23 | |
| Db | 110 | CTTTTCATCCCACTGCCTTTGA 132 | |
| RESULT 9 | | | |
| CD067162 | | | |
| LOCUS | MA1-0049T-R073-A09-U.G | MA1-0049 Schistosoma mansoni | EST 14-SEP-2003 |
| DEFINITION | MA1-0049T-R073-A09.G, mRNA sequence. | | |
| ACCESSION | CD067162 | | |
| VERSION | CD067162.1 | GI:34618281 | |
| KEYWORDS | EST. | | |
| SOURCE | Schistosoma mansoni | | |
| ORGANISM | Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomidae; Schistosoma. | | |
| REFERENCE | 1 (bases 1 to 228) Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M., Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr., Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F., Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L., Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A., Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A., Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T., Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M., Setubal,J.C., Leite,L.C.C. and Dias-Neto,E. | | |
| AUTHORS | Transcriptome analysis of the acelomate human parasite Schistosoma mansoni | | |
| TITLE | Nat. Genet. 35 (2), 148-157 (2003) | | |
| JOURNAL | 12973350 | | |
| MEDLINE | 22879926 | | |
| PUBMED | 12973350 | | |
| COMMENT | Contact: Dr. Sergio Verjovski-Almeida Departamento de Bioquímica Instituto de Química - Universidade de São Paulo Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP, Brasil Tel: +55-11-3091-2173 Fax: +55-11-3091-2186 Email: verjo@iq.usp.br This sequence was derived from the PAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL http://bioinfo.iq.usp.br/schisto/ Plate: MA1-0049T-R073 row: 9 column: A. | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..228 /organism="Schistosoma mansoni" /mol_type="mRNA" /db_xref="taxon:6183" /clone="MA1-0049T-R073-A09.G" | | |

```
/sex="mixed pool"
/dev_stage="adult"
/lab_host="Mesocricetus auratus"
/clone_lib="MAL-0049"
```

ORIGIN

```
Query Match 79.1%; Score 18.2; DB 6; Length 228;
Best Local Similarity 87.0%; Pred. No. 8.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 CCTTTTCATCCAACTACCACTGA 23
```

```
Db 103 CCTTCTCCTCCAAATACCACTGA 125
```

RESULT 10

BB486889

```
LOCUS BB486889 281 bp mRNA linear EST 23-JUL-2000
DEFINITION BB486889 RIKEN full-length enriched, 13 days embryo lung Mus
musculus cDNA clone D430039B11 3' similar to U49351 Mus musculus
lysosomal alpha-glucosidase mRNA, mRNA sequence.
```

ACCESSION

BB486889

BB486889.1 GI:9404498

VERSION

EST.

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus

ORGANISM

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 281)

AUTHORS

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, M., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toyota, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

TITLE

Yoshihide Hayashizaki

JOURNAL

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,

Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Oza, Y., Muramatsu, M.,

Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

FEATURES

Location/Qualifiers

1..281

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="D430039B11"

```
/tissue_type="lung"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 13 days embryo
lung"
```

/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'

GAGAGAGAGCGCGCAACTCGAGTTTTTTTTTTTTTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'

GAGAGAGATTCGAGTTAATTAATATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

```
Query Match 79.1%; Score 18.2; DB 2; Length 281;
Best Local Similarity 87.0%; Pred. No. 8.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 CCTTTTCATCCAACTACCACTGA 23
```

```
Db 16 CCTTTTCATCCATCCACCCCTGA 38
```

RESULT 11

AV107656

LOCUS

DEFINITION

AV107656

ACCESSION

AV107656

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 295)

REFERENCE

AUTHORS

Carninci, P., Shibata, K., Oza, Y., Konno, H., Itoh, M., Aizawa, K.,

Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,

Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,

Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H.,

Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,

Sugahara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y.,

Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,

Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp

Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

FEATURES

Location/Qualifiers

1..295

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

```

/db_xref="taxon:10090"
/clone="2510040A06"
/sex="mixed"
/tissue_type="liver"
/dev_stage="13-day embryo"
/clone_lib="Mus musculus liver C57BL/6J 13-day embryo"

ORIGIN

```

```

Query Match      79.1%; Score 18.2; DB 1; Length 295;
Best Local Similarity 87.0%; Pred. No. 8.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 CCTTTTCATCCCACTACCACTGA 23
    |||||
Db 33 CCTTTTCATCCCACTACCACTGA 55

```

```

RESULT 12
BH170823
LOCUS      394 bp      DNA      linear      GSS 03-OCT-2001
DEFINITION SALK_003398 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_003398, genomic survey sequence.
ACCESSION  BH170823
VERSION    BH170823.1 GI:15906526
KEYWORDS   GSS.
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana

```

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 394)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.O., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (IGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.

```

```

FEATURES
source
Location/Qualifiers
1..394
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_003398"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

```

```

ORIGIN
Query Match      79.1%; Score 18.2; DB 8; Length 394;
Best Local Similarity 87.0%; Pred. No. 9.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 CCTTTTCATCCCACTACCACTGA 23
    |||||
Db 315 CATTTTATCCCACTACCACTAA 337

```

```

RESULT 13

```

```

CB101123/C
LOCUS      452 bp      mRNA      linear      EST 28-JAN-2003
DEFINITION K16609.y1 Ascaris suum embryo pAMP1 v2 Ascaris suum cDNA 5', mRNA
            sequence.
ACCESSION  CB101123
VERSION    CB101123.1 GI:27926930
KEYWORDS   EST.
SOURCE     Ascaris suum (pig roundworm)
ORGANISM   Ascaris suum

```

```

Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida;
Ascaridoidea; Ascaridae; Ascaris.
1 (bases 1 to 452)
McCarte,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
Bowers,Y., Gibbons,M., Riter,E., Bennett,J., Franklin,C.,
Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Claire Murphy and Dr. James McCarter
at Washington University, St. Louis. The cDNA was made by using
Dynabead oligo-dT priming (Dyna), PCR based library using a
modified protocol from the SMART PCR cDNA Synthesis Kit from
Clontech. Directionally cloned into the UDG sites of pAMP1. 30-60
cell embryo material was provided by Dr. Richard Davis of City
University of New York Graduate Center, College of Staten Island,
Staten Island, NY (redavis@postbox.csi.cuny.edu).
Seq primer: -40RP from Gibco
High quality sequence stop: 422.

```

```

FEATURES
source
Location/Qualifiers
1..452
/organism="Ascaris suum"
/mol_type="mRNA"
/db_xref="taxon:6253"
/dev_stage="30-60 cell embryo"
/clone_lib="Ascaris suum embryo pAMP1 v2"
/clone_host="DH10B Vector"
/notes="Vector: pAMP1; Site 1: NotI; Site 2: SalI; The
library was constructed by Claire Murphy and Dr. James
McCarter at Washington University, St. Louis. The cDNA was
made by using Dynabead oligo-dT priming (Dyna). PCR based
library using a modified protocol from the SMART PCR cDNA
Synthesis Kit from Clontech. Directionally cloned into the
UDG sites of pAMP1. 30-60 cell embryo material was
provided by Dr. Richard Davis of City University of New
York Graduate Center, College of Staten Island, Staten
Island, NY (redavis@postbox.csi.cuny.edu)."

```

```

ORIGIN
Query Match      79.1%; Score 18.2; DB 6; Length 452;
Best Local Similarity 87.0%; Pred. No. 9.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 CCTTTTCATCCCACTACCACTGA 23
    |||||
Db 25 CCTTGTCATCGAATCATCACTGA 3

```

```

RESULT 14
CE706312/C
LOCUS      581 bp      DNA      linear      GSS 29-SEP-2003
DEFINITION tigr-gss-dog-17000369184273 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION  CE706312

```



```
VERSION      CE706312.1  GI:37025703
KEYWORDS     GSS.
SOURCE       Canis familiaris (dog)
ORGANISM     Canis familiaris
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE    Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
AUTHORS      1 (bases 1 to 581)
             Kirkness, E.F., Bakna, V., Halpern, A.L., Levy, S., Remington, K.,
             Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
             Venter, J.C.
TITLE        The dog genome: survey sequencing and comparative analysis
JOURNAL      Science 301 (5641), 1898-1903 (2003)
MEDLINE      22875432
PUBMED       14512627
COMMENT      Contact: Kirkness EF
             The Institute for Genomic Research
             Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
             Rockville, MD 20850, USA
             Tel: 301-838-0200
             Fax: 301-838-0208
             Email: ekirknes@tigr.org
             Class: shotgun.
FEATURES     Location/Qualifiers
             1..581
                /organism="Canis familiaris"
                /mol_type="genomic DNA"
                /strain="Standard Poodle"
                /db_xref="taxon:9615"
                /clone_lib="Dog Library"
                /note="Site 1: BstXI; Libraries were prepared from
                peripheral blood"
ORIGIN
Query Match      79.1%; Score 18.2; DB 9; Length 581;
Best Local Similarity 87.0%; Pred. No. 9.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CCTTTTCATCCCACTACCACTGA 23
        |||||
Db      389  CCTTTTCATCCCACTACCACTGA 367

RESULT 15
LOCUS   CC320497               1184 bp    DNA    linear    GSS 14-MAY-2003
DEFINITION
TAM32-27E3_Sp6.1 TAM32 Gallus gallus genomic clone TAM32-27E3,
genomic survey sequence.
ACCESSION CC320497
VERSION   CC320497.1  GI:30714555
KEYWORDS  GSS.
SOURCE    Gallus gallus (chicken)
ORGANISM  Gallus gallus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
          Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1184)
          Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
          Warren, W., Graves, T., Mardis, E. and Wilson, R.
          Gallus gallus BAC End Reads
          Unpublished (2003)
          Contact: Richard K. Wilson
          Genome Sequencing Center
          Washington University School of Medicine
          Email: submission@watson.wustl.edu
          Insert Length: 152000 Std Error: 0.00
          Seq primer: Sp6 ATTAGTGACACTATAG
          Class: BAC ends
          High quality sequence start: 8
          High quality sequence stop: 751.
FEATURES     Location/Qualifiers
             1..1184
                /organism="Gallus gallus"
                /mol_type="genomic DNA"
```

```
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="TAM32-27E3"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="TAM32"
/note="Vector: pECBAC1; Site 1: EcoRI; Site 2: EcoRI;
TAM32 Female Chicken library - for library and clone
ordering information: http://www.hbz.tamu.edu"
ORIGIN
Query Match      79.1%; Score 18.2; DB 8; Length 1184;
Best Local Similarity 87.0%; Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CCTTTTCATCCCACTACCACTGA 23
        |||||
Db      185  CATTCTCATCCCACTACCACTAA 207

Search completed: August 13, 2005, 06:44:57
Job time : 1486.12 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 01:08:33 ; Search time 788.298 Seconds
(without alignments)
1536.704 Million cell updates/sec

Title: US-10-673-854-3

Perfect score: 25
Sequence: 1 tgaagaatttcagttcatagcttgt 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 25 | 100.0 | 142099 | 9 | AC002509 |
| 2 | 25 | 100.0 | 176872 | 9 | AC012078 |
| 3 | 25 | 100.0 | 212280 | 9 | HSB362E11 |
| 4 | 20.2 | 80.8 | 144552 | 8 | CNS08CEK |
| 5 | 20.2 | 80.8 | 155327 | 9 | AC138761 |
| 6 | 20.2 | 80.8 | 157038 | 2 | EX957355 |
| 7 | 20.2 | 80.8 | 167472 | 5 | EX001041 |
| 8 | 20.2 | 80.8 | 198756 | 9 | AC087575 |
| 9 | 20.2 | 80.8 | 199035 | 2 | AC022263 |
| 10 | 20.2 | 80.8 | 204512 | 2 | AC119029 |
| 11 | 20.2 | 80.8 | 215185 | 2 | EX927073 |
| 12 | 20.2 | 80.8 | 226228 | 2 | AC098027 |
| 13 | 19.8 | 79.2 | 44848 | 6 | AR438843 |
| 14 | 19.8 | 79.2 | 57633 | 9 | AC093382 |
| 15 | 19.8 | 79.2 | 72040 | 2 | AC016347 |
| 16 | 19.8 | 79.2 | 81017 | 2 | AC025010 |
| 17 | 19.8 | 79.2 | 91798 | 9 | HSXB152G3 |
| 18 | 19.8 | 79.2 | 108418 | 9 | AC007001 |
| 19 | 19.8 | 79.2 | 113587 | 9 | AL590702 |

| | | | | | | | |
|---|----|------|------|--------|----|----------|---------------------|
| c | 20 | 19.8 | 79.2 | 119024 | 9 | AL159987 | AL159987 Human DNA |
| c | 21 | 19.8 | 79.2 | 124457 | 9 | AC115115 | AC115115 Homo sapi |
| c | 22 | 19.8 | 79.2 | 137246 | 9 | HS49J10 | Z84572 Human DNA s |
| c | 23 | 19.8 | 79.2 | 145859 | 9 | AC009510 | AC009510 Homo sapi |
| c | 24 | 19.8 | 79.2 | 149188 | 9 | AC114781 | AC114781 Homo sapi |
| c | 25 | 19.8 | 79.2 | 149403 | 2 | AC130461 | AC130461 Homo sapi |
| c | 26 | 19.8 | 79.2 | 153468 | 9 | AC113208 | AC113208 Homo sapi |
| c | 27 | 19.8 | 79.2 | 154439 | 2 | AC073161 | AC073161 Homo sapi |
| c | 28 | 19.8 | 79.2 | 157267 | 9 | AC008011 | AC008011 Homo sapi |
| c | 29 | 19.8 | 79.2 | 163777 | 9 | AC067932 | AC067932 Homo sapi |
| c | 30 | 19.8 | 79.2 | 164858 | 9 | AC026634 | AC026634 Homo sapi |
| c | 31 | 19.8 | 79.2 | 165558 | 4 | AP006185 | AP006185 Sus scrofa |
| c | 32 | 19.8 | 79.2 | 167476 | 9 | AC116038 | AC116038 Homo sapi |
| c | 33 | 19.8 | 79.2 | 167580 | 2 | AC146124 | AC146124 Pan trogl |
| c | 34 | 19.8 | 79.2 | 168487 | 9 | AL138820 | AL138820 Human DNA |
| c | 35 | 19.8 | 79.2 | 169416 | 9 | AC022078 | AC022078 Homo sapi |
| c | 36 | 19.8 | 79.2 | 174365 | 2 | AC024552 | AC024552 Homo sapi |
| c | 37 | 19.8 | 79.2 | 179755 | 9 | AL603831 | AL603831 Human DNA |
| c | 38 | 19.8 | 79.2 | 180443 | 2 | AC148366 | AC148366 Callithri |
| c | 39 | 19.8 | 79.2 | 184396 | 9 | AC010853 | AC010853 Homo sapi |
| c | 40 | 19.8 | 79.2 | 184518 | 9 | AC099524 | AC099524 Homo sapi |
| c | 41 | 19.8 | 79.2 | 188032 | 2 | AC016393 | AC016393 Homo sapi |
| c | 42 | 19.8 | 79.2 | 189590 | 5 | EX470167 | EX470167 Zebrafish |
| c | 43 | 19.8 | 79.2 | 196477 | 10 | AC098390 | AC098390 Rattus no |
| c | 44 | 19.8 | 79.2 | 202950 | 9 | AC017100 | AC017100 Homo sapi |
| c | 45 | 19.8 | 79.2 | 203270 | 9 | AC064807 | AC064807 Homo sapi |

ALIGNMENTS

| | | | | | |
|------------|---|---|-----|--------|-----------------|
| RESULT 1 | AC002509 | 142099 bp | DNA | linear | PRI 25-NOV-1998 |
| LOCUS | AC002509 | Homo sapiens chromosome Y, clone 2Y, complete sequence. | | | |
| DEFINITION | AC002509 | Homo sapiens chromosome Y, clone 2Y, complete sequence. | | | |
| ACCESSION | AC002509.1 | GI:3927860 | | | |
| VERSION | HTG. | | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | HTG. | | | | |
| ORGANISM | Homo sapiens (human) | | | | |
| | Homo sapiens | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 142099) | | | | |
| AUTHORS | Birren,B., Linton,L., Nusbaum,C. and Lander,E. | | | | |
| TITLE | Homo sapiens chromosome Y, clone 2Y | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 142099) | | | | |
| AUTHORS | Hawkins,T.L., Birren,B.W., Fauman,K.H., Nusbaum,C., Lander,E.S., McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P., Daly,M.J., Devon,K., Dewar,K., Forrest,C., Gage,D., Geraghty,K., Hagos,B., Huang,J., Hui,L., Jacotot,L., Kirby,A., Lane,M., MacKenzie,J., Marquis,N., McDermott,J., Molla,M., Morrow,J., Nachman,A., Naylor,J., Nusbaum,C., O'Connor,T., Olotu,A., Peterson,K., Reeve,M.P., Roberts,D., Rollins,G., Stillewell,J., Stone,C., Strickland,C., Sydney,K., Tang,L., Wilmer,F., Zemtseva,I., and Zody,M. | | | | |
| | Direct Submission | | | | |
| | Submitted (27-AUG-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA | | | | |
| JOURNAL | Research, 320 Charles Street, Cambridge, MA 02141, USA | | | | |
| REFERENCE | 3 (bases 1 to 142099) | | | | |
| AUTHORS | Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donisian,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,K., Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McSwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Roy,A., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Robertson,D., Roy,A., Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Tichovolsky,N., Torruella-Miller,I., | | | | |

Vassiliev,H., Vo.A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
 Ye,W.J., Zhao,J. and Zody,M.
 Direct Submission
 Submitted (25-NOV-1998) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Nov 25, 1998 this sequence version replaced gi:3924665.
 All repeats were identified using RepeatMasker; Smit, A.F.A. &
 Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.

FEATURES

Location/Qualifiers

1. .142099

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="Y"

/map="Y"

/clone="2Y"

/clone_lib="unknown"

complement(315. .438)

/rpt_family="MER61-internal"

complement(2160. .2521)

/rpt_family="MER50"

2522. .2578

/rpt_family="AluSg/x"

2859. .2919

/rpt_family="MER51-internal"

complement(2920. .2978)

/rpt_family="(CA)n"

2990. .3699

/rpt_family="MER4-internal"

complement(3700. .3814)

/rpt_family="LTR8"

3815. .4468

/rpt_family="MER4-internal"

4523. .7011

/rpt_family="MER4-internal"

complement(7295. .7598)

/rpt_family="AluSp"

9162. .9349

/rpt_family="MER4-internal"

9361. .9509

/rpt_family="MER4A2"

9510. .9811

/rpt_family="MER4A"

complement(9812. .9962)

/rpt_family="LIM1"

complement(10212. .10830)

/rpt_family="LIM4"

complement(10831. .11120)

/rpt_family="AluY"

complement(11121. .11337)

/rpt_family="LIM4"

complement(11508. .12117)

/rpt_family="LIMEC"

complement(12315. .12626)

/rpt_family="LIME"

complement(12923. .13136)

/rpt_family="MIR"

complement(13246. .13328)

/note="Single-stranded terminator coverage."

13761. .13782

/rpt_family="AT_rich"

14930. .14950

/rpt_family="AT_rich"

15128. .15231

/rpt_family="HAL1"

15292. .15552

/rpt_family="HAL1"

complement(15690. .16275)

/rpt_family="MER4B"

complement(16397. .16551)

/rpt_family="MER66-internal"

16825. .16846

repeat_region
 complement(17272. .17680)
 /rpt_family="MER4-internal"
 repeat_region
 complement(17687. .17920)
 /rpt_family="MER31-internal"
 repeat_region
 complement(18046. .18789)
 /rpt_family="MER50"
 repeat_region
 complement(19887. .20110)
 /rpt_family="MER57-internal"
 repeat_region
 complement(20133. .20299)
 /rpt_family="MER4-internal"
 repeat_region
 20300. .20593
 /rpt_family="AluY"
 repeat_region
 complement(20594. .20673)
 /rpt_family="MER4-internal"
 repeat_region
 complement(20674. .20974)
 /rpt_family="AluSg"
 repeat_region
 complement(20975. .21867)
 /rpt_family="MER4-internal"
 repeat_region
 complement(21372. .21498)
 /rpt_family="MER51-internal"
 repeat_region
 complement(21499. .21804)
 /rpt_family="AluSg"
 repeat_region
 complement(21805. .21964)
 /rpt_family="MER51-internal"
 repeat_region
 complement(21965. .22251)
 /rpt_family="MER4-internal"
 repeat_region
 complement(22252. .22441)
 /rpt_family="MER51-internal"
 repeat_region
 complement(22490. .22671)
 /rpt_family="MER83-internal"
 repeat_region
 complement(22589. .22693)
 /rpt_family="MER65-internal"
 repeat_region
 23324. .23353
 /rpt_family="AT_rich"
 repeat_region
 complement(23437. .23977)
 /rpt_family="MER4B"
 repeat_region
 complement(23978. .24177)
 /rpt_family="MLT2A"
 repeat_region
 24557. .24891
 /rpt_family="HAL1"
 repeat_region
 25476. .25758
 /rpt_family="L2"
 repeat_region
 26069. .26101
 /rpt_family="AT_rich"
 repeat_region
 26265. .26489
 /rpt_family="MLTID"
 repeat_region
 26490. .26517
 /rpt_family="(CAAA)n"
 repeat_region
 26518. .26780
 /rpt_family="MLTID"
 repeat_region
 27156. .27180
 /rpt_family="AT_rich"
 repeat_region
 complement(27417. .27797)
 /rpt_family="L2"
 repeat_region
 complement(27963. .28332)
 /rpt_family="MSTA"
 repeat_region
 complement(28333. .29268)
 /rpt_family="MSTA-internal"
 repeat_region
 29269. .29580
 /rpt_family="AluY"
 repeat_region
 complement(29581. .30301)
 /rpt_family="MSTA-internal"
 repeat_region
 complement(30304. .30679)
 /rpt_family="MSTA"
 repeat_region
 complement(30680. .31159)
 /rpt_family="LTR40a"
 repeat_region
 complement(31325. .31655)
 /rpt_family="LIME"
 repeat_region
 31692. .31731
 /rpt_family="AT_rich"
 repeat_region
 31803. .31927
 /rpt_family="AT_rich"

```

repeat_region complement(32107..33826)
repeat_region /rpt_family="L1M4"
repeat_region 33831..34538
repeat_region /rpt_family="L1ME3"
repeat_region 34987..35012
repeat_region /rpt_family="AT_rich"
repeat_region 35108..35669
repeat_region /rpt_family="MER4B"
repeat_region 36280..36322
repeat_region /rpt_family="AT_rich"
repeat_region complement(36385..36519)
repeat_region /rpt_family="MIR"
repeat_region complement(36755..37203)
repeat_region /rpt_family="MER4C"

Query Match 100.0%; Score 25; DB 9; Length 142099;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGAAGAAATTCAGTTCATAGCTTGT 25
|||||
Db 35055 TGAAGAAATTCAGTTCATAGCTTGT 35079
|||||

RESULT 2
AC012078 176872 bp DNA linear PRI 30-SEP-2000
LOCUS Homo sapiens BAC clone RP11-539022 from Y, complete sequence.
AC012078
ACCESSION AC012078.3 GI:7684580
VERSION
KEYWORDS HMG.
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS Hou, S., Maupin, R. and Gibson, A.
TITLE The sequence of Homo sapiens BAC clone RP11-539022
JOURNAL Unpublished
JOURNAL 3 (bases 1 to 176872)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 176872)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 176872)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 176872)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 3, 2000 this sequence version replaced gi:7631053.
-----
COMMENT Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc

```

Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0539022

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skalecky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Ooegawa, K., Woon, P.Y., Zhao, B., Frengen, B., Tateno, M., Catanesse, J.J. and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-530K5; the clone sequenced to the right is RP11-33605. Actual start of this clone is at base position 1 of RP11-539022; actual end is at base position 176872 of RP11-539022.

| FEATURES | Location/Qualifiers |
|---------------|--------------------------|
| source | 1..176872 |
| | /organism="Homo sapiens" |
| | /mol_type="genomic DNA" |
| | /db_xref="taxon:9606" |
| | /chromosome="Y" |
| | /map="Y" |
| repeat_region | /clone="RP11-539022" |
| | /clone_lib="RP11-11" |
| | 1..204 |
| | /rpt_family="L1" |
| repeat_region | 202..1293 |
| | /rpt_family="L1" |
| repeat_region | 1294..1324 |
| | /rpt_family="(CAAA)n" |
| repeat_region | 1325..2158 |
| | /rpt_family="L1" |
| repeat_region | 2159..2339 |
| | /rpt_family="L1" |
| repeat_region | 2340..3099 |
| | /rpt_family="L1" |
| repeat_region | 3120..3219 |
| | /rpt_family="Alu" |
| repeat_region | 3220..3945 |
| | /rpt_family="L1" |
| repeat_region | 3945..4535 |
| | /rpt_family="L1" |
| repeat_region | 4565..4619 |
| | /rpt_family="Alu" |
| repeat_region | 4638..5206 |
| | /rpt_family="L1" |
| repeat_region | 5255..5268 |


```

/note="MER34 repeat: matches 498. .534 of consensus"
120. .403
/note="MER39b repeat: matches 68. .453 of consensus"
404. .498
/note="MER39 repeat: matches 7. .103 of consensus"
502. .1573
/note="LIM4 repeat: matches 4008. .5095 of consensus"
1569. .1914
/note="LIM4 repeat: matches 3525. .3878 of consensus"
1906. .2066
/note="LIMD repeat: matches 1376. .1528 of consensus"
2112. .2378
/note="ALUjb repeat: matches 5. .274 of consensus"
2379. .2635
/note="LIMD repeat: matches 1126. .1402 of consensus"
2651. .2697
/note="MLT2D repeat: matches 504. .550 of consensus"
2840. .3252
/note="MLT2D repeat: matches 77. .514 of consensus"
3351. .3420
/note="MLT2D repeat: matches 1. .71 of consensus"
3421. .3902
/note="HERVL repeat: matches 5231. .5757 of consensus"
3903. .4462
/note="280 copies 2 mer tt 57% conserved"
4465. .4782
/note="ALUjo repeat: matches 1. .312 of consensus"
4783. .6107
/note="HERVL repeat: matches 3896. .5242 of consensus"
6238. .6378
/note="LIMD repeat: matches 973. .1115 of consensus"
6640. .6783
/note="LIMC5 repeat: matches 7751. .7910 of consensus"
7282. .7567
/note="MLT1H repeat: matches 1. .302 of consensus"
complement (7766. .8138)
/note="match: GSS: Em:AQ098528"
complement (7769. .8123)
/note="match: GSS: Em:AQ105610"
complement (7779. .8205)
/note="match: GSS: Em:AQ239326"
8002. .8469
/note="MLT1D repeat: matches 1. .503 of consensus"
8477. .8654
/note="MIR repeat: matches 75. .256 of consensus"
8715. .8812
/note="49 copies 2 mer aa 61% conserved"
complement (9391. .9893)
/genes="BB362B11.1"
/pseudo
complement (9391. .9893)
/genes="BB362B11.1"
/note="B362B11.1"
/note="B362B11.1 (pseudogene similar to mouse GSG-154 and
mosquito MRRG)
match: cDNAs: Em:AK001774
match: proteins: Sw:P50636 Tr:Q9Y4Y1 Tr:P90662 Tr:P91067"
/pseudo
/codon_start=1
/evidence=not_experimental
9919. .10251
/note="match: STS: Em:T15642"
10912. .13011
/note="L1 repeat: matches 2373. .4514 of consensus"
13018. .13758
/note="TIGER1 repeat: matches 1. .774 of consensus"
13759. .13814
/note="LIM4c repeat: matches 1557. .1609 of consensus"
13815. .14766
/note="LITR5 repeat: matches 1. .969 of consensus"
14767. .14882
/note="LIM4c repeat: matches 1609. .1737 of consensus"
15152. .15106
/note="LIM3 repeat: matches 4670. .5642 of consensus"

```

```

repeat_region
16107. .16560
/note="LITR46 repeat: matches 1. .461 of consensus"
18749. .19851
/note="HERVH repeat: matches 3994. .5500 of consensus"
19925. .22648
/note="HERVH21 repeat: matches 1275. .3767 of consensus"
complement (22988. .23428)
/note="match: GSS: Em:AQ619480"
23057. .23492
/note="match: GSS: Em:AQ833119"
23749. .24192
/note="LITR46 repeat: matches 1. .461 of consensus"
24194. .24348
/note="LIM3 repeat: matches 5639. .5793 of consensus"
24386. .24485
/note="LIM9 repeat: matches 5995. .6096 of consensus"
24497. .25002
/note="LIM9 repeat: matches 5794. .6308 of consensus"
complement (25454. .26350)
/note="match: GSS: Em:AQ894184"
25505. .25526
/note="11 copies 2 mer ta 100% conserved"
25587. .25823
/note="MIR repeat: matches 17. .262 of consensus"
complement (25882. .26240)
/note="match: GSS: Em:AQ929805"
25885. .26180
/note="match: GSS: Em:AQ754783"
complement (25910. .26219)
/note="match: GSS: Em:AQ785533"
25945. .26256
/note="match: GSS: Em:AQ628742"
25949. .26676
/note="match: GSS: Em:ALI08886"
25950. .26147
/note="match: GSS: Em:AQ230662"
25951. .26260
/note="match: GSS: Em:AQ866229"
complement (25951. .26227)
/note="match: GSS: Em:AQ541753"
complement (25951. .26105)
/note="match: GSS: Em:AQ017857"
25951. .25978
/note="14 copies 2 mer ta 92% conserved"
25954. .26352
/note="match: GSS: Em:AQ525540"
complement (25978. .26215)
/note="match: GSS: Em:AQ023352"
complement (25979. .26131)
/note="match: GSS: Em:AQ322892"
25980. .26025
/note="23 copies 2 mer ta 84% conserved"
complement (25984. .26583)
/note="match: GSS: Em:AQ321139"
complement (25988. .26185)
/note="match: GSS: Em:AQ226794"
25994. .26263
/note="match: GSS: Em:AZ018667 Em:AZ052220"
complement (25998. .26169)
/note="match: STS: Em:AU047022"
complement (26000. .26305)
/note="match: GSS: Em:AQ123830"
26005. .26229
/note="match: GSS: Em:AZ007683"
26009. .26160
/note="match: STS: Em:HSB017WB9"
26009. .26147
/note="match: GSS: Em:AZ069301"
26037. .26181
/note="match: GSS: Em:AQ871738"
complement (26046. .26369)
/note="match: GSS: Em:AQ827142"
26046. .26360

```

Query Match 100.0%; Score 25; DB 9; Length 212280;
 Best Local Similarity 100.0%; Pred. No. 1.6; Mismatches 0; Indels 0; Gaps 0;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCTTGT 25
 |||||
 Db 116062 TGAAGAAATTCAGTTCATAGCTTGT 116086
 |||||

RESULT 4
 CNS08CBK/c
 LOCUS
 DEFINITION Oryza sativa chromosome 12, BAC OSJNBa0017A21 of library OSJNBa
 from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza
 sativa (rice), complete sequence.

ACCESSION AL928747
 VERSION AL928747.3 GI:28412540
 KEYWORDS HTG.

ORGANISM Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 144552)

AUTHORS Choisne,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
 Segurenbach,J. and Quetier,F.

TITLE Oryza sativa chromosome 12 sequencing

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 144552)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT On Feb 18, 2003 this sequence version replaced gi:24430303.

Center: Genoscope / Centre National de Sequencage

Center code: GS

Web site: http://www.genoscope.cns.fr/

Contact: SeqRef@genoscope.cns.fr

 The following sequence is oriented from the T7 to the SP6 end. The
 nucleotide sequence of this BAC clone was generated by combining
 Syngenta and Genoscope sequencing data.
 Upstream BAC (overlapping the T7 end) : P0185F07 (AC=BX667525)
 Downstream BAC (overlapping the SP6 end) : OJ1111_F12 (AC=AL732536)

 FINISHING BOUNDARIES
 FINISHED SEGMENT STARTS AT BASE 1
 FINISHED SEGMENT ENDS AT BASE 110216

FEATURES Location/Qualifiers

1. .144552
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:39947"
 /chromosome="12"
 /clone="OSJNBa0017A21"
 /clone_lib="OSJNBa"

ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 144552;
 Best Local Similarity 88.0%; Pred. No. 2e+02; Mismatches 0; Indels 3; Gaps 0;
 Matches 22; Conservative 0; Mismatches 0; Indels 3; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCTTGT 25
 |||||
 Db 66142 TGAAGAAATTCAGTTCATAGCTTGT 66118
 |||||

RESULT 5.

AC138761/c
 LOCUS
 DEFINITION Homo sapiens
 ACCESSION AC138761
 VERSION AC138761.4
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 155327)

AUTHORS Birren,B., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 17, clone RP11-1109M24

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 155327)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
 Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kanat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
 Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
 Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
 Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 155327)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collamore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fero,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kanat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (13-FEB-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 155327)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collamore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fero,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kanat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,

Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramaamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (01-MAR-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 1, 2003 this sequence version replaced gi:28301978.
 All repeats were identified using RepeatMasker:
 Smith, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L29291
 Center clone name: 1109_M_24

FEATURES

Location/Qualifiers

Source
 1. .155327
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="RP11-1109M24"
 /clone_lib="RPC1-11 Human Male BAC"
 17..72
 /rpt_family="CT-rich"
 72..145
 /rpt_family="(TA)n"
 complement(291..500)
 /rpt_family="MIR"
 527..568
 /rpt_family="AT-rich"
 3043..3079
 /rpt_family="U2"
 3307..3333
 /rpt_family="(TTTTTG)n"
 3543..3849
 /rpt_family="AluY"
 4155..4182
 /rpt_family="AT-rich"
 4757..4780
 /rpt_family="AT-rich"
 5591..5653
 /rpt_family="(TATATC)n"
 complement(7305..7355)
 /rpt_family="L1M4"
 complement(7485..8409)
 /rpt_family="L1MDa"
 9060..9088
 /rpt_family="(TC)n"
 complement(9740..9897)
 /rpt_family="MIR"
 10340..10370
 /rpt_family="(TTCA)n"
 complement(10371..10522)
 /rpt_family="AluSg/x"
 complement(10831..11204)
 /rpt_family="THE1C"
 complement(11205..12793)
 /rpt_family="THE1C-int"
 complement(12794..13177)
 /rpt_family="THE1C"
 13179..13980
 /rpt_family="L2"

repeat_region 14702..15028
 /rpt_family="MER2"
 repeat_region 15087..15562
 /rpt_family="MER70A"
 repeat_region 15761..15933
 /rpt_family="L1ME3B"
 repeat_region 15967..16422
 /rpt_family="L1ME3B"
 repeat_region 16436..16459
 /rpt_family="AT-rich"
 repeat_region 16712..17467
 /rpt_family="L1ME3B"
 repeat_region complement(17468..17947)
 /rpt_family="MLT1D"
 repeat_region 17948..18034
 /rpt_family="L1ME3B"
 repeat_region 18747..18836
 /rpt_family="AT-rich"
 repeat_region complement(19087..19362)
 /rpt_family="AluX"
 repeat_region complement(19585..19878)
 /rpt_family="AluY"
 repeat_region complement(19890..20046)
 /rpt_family="AluSp"
 repeat_region complement(22092..22373)
 /rpt_family="AluX"
 repeat_region 23331..23384
 /rpt_family="AT-rich"
 repeat_region complement(23597..23644)
 /rpt_family="LTR66"
 repeat_region 24249..24549
 /rpt_family="AluSp"
 repeat_region 24778..24804
 /rpt_family="(CA)n"
 repeat_region 24855..25198
 /rpt_family="THE1B"
 repeat_region 25208..26788
 /rpt_family="THE1B-int"
 repeat_region 26790..26993
 /rpt_family="THE1B"
 repeat_region 27626..27661
 /rpt_family="AT-rich"
 repeat_region 28903..28931
 /rpt_family="AT-rich"
 repeat_region complement(29301..29597)
 /rpt_family="AluSc"
 repeat_region complement(29687..29914)
 /rpt_family="L1MA3"
 repeat_region 30315..30352
 /rpt_family="AT-rich"
 repeat_region 30632..30682
 /rpt_family="AT-rich"
 repeat_region complement(30713..31593)
 /rpt_family="L1MA3"
 repeat_region 31594..32226

Query Match 80.8%; Score 20.2; DB 9; Length 155327;
 Best Local Similarity 88.0%; Pred.No.2e+02; 3; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 0;

QY 1 TGAAGAAATTCAGTTCATAGCTTGT 25
 |||||
 Db 72347 TGAAGAAATTCAGTTCATAGCTTTT 72323

RESULT 6

BX957355
 LOCUS BX957355 157038 bp DNA linear HTG 01-MAR-2004
 DEFINITION Danio rerio clone CH211-66115, *** SEQUENCING IN PROGRESS ***, 8
 unorderd pieces.
 ACCESSION BX957355
 VERSION BX957355.3 GI:44845041
 KEYWORDS HTG; HTGS_PHASE1.


```

/rpt_family="AluSg"
complement(5197..5481)
/rpt_family="AluSx"
complement(5707..5838)
/rpt_family="FLAM_C"
6298..6597
/rpt_family="AluJb"
6600..7279
/rpt_family="Tigger1"
complement(7280..7570)
/rpt_family="AluSp"
7628..7821
/rpt_family="MER115"
8003..8103
/rpt_family="TTATA)n"
8125..8626
/rpt_family="Tigger1"
complement(8627..8904)
/rpt_family="AluSx"
8905..9037
/rpt_family="Tigger1"
complement(9041..9331)
/rpt_family="LIR16C"
9518..9606
/rpt_family="MER115"
9723..10065
/rpt_family="MER115"
complement(10467..10750)
/rpt_family="L1M44"
complement(11038..11252)
/rpt_family="MIR"
11357..11671
/rpt_family="AluSx"
complement(11766..11866)
/rpt_family="L2"
complement(12415..12629)
/rpt_family="MIR"
12775..12805
/rpt_family="AT_rich"
13003..13046
/rpt_family="AT_rich"
complement(13289..13376)
/rpt_family="L1PA16"
13617..13754
/rpt_family="FLAM_A"
complement(13756..14094)
/rpt_family="MER93"
14538..14558
/rpt_family="AT_rich"
14571..15059
/rpt_family="L1MB3"
15060..15383
/rpt_family="AluSg"
15384..15959
/rpt_family="L1MB3"
15968..16030
/rpt_family="AT_rich"
16512..16538
/rpt_family="AT_rich"
complement(16541..16839)
/rpt_family="AluY"
17404..17700
/rpt_family="AluSg"
17717..17993
/rpt_family="AluSp"
17995..18041
/rpt_family="TAAAA)n"
complement(18048..18322)
/rpt_family="L1MD1"
complement(18493..18674)
/rpt_family="AluSg/x"
18800..19089
/rpt_family="AluJb"

```

```

* 83001 83100: gap of unknown length
* 83101 118000: contig of 34900 bp in length
* 118001 118100: gap of unknown length
* 118101 199035: contig of 80935 bp in length.
FEATURES
    source      Location/Qualifiers
                1..199035
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="UL"
                /clone="RP11-305L6"
ORIGIN
Query Match      80.8%; Score 20.2; DB 2; Length 199035;
Best Local Similarity 88.0%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAAGAATTTCAGTTCATAGCTTGT 25
Db 30469 TGAAGAATTTCAGTTCATAGCTTTT 30493

RESULT 10
AC119029/c
LOCUS      AC119029      204512 bp      DNA      linear      HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-50113, *** SEQUENCING IN PROGRESS
***
AC119029
AC119029.4 GI:25137856
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE
AUTHORS     Muzny, D., Marie, E., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,
            Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
            Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
            Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P.,
            Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
            Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
            Cardenas, J., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
            Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
            Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
            Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
            Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
            Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
            Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
            Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
            Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
            Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
            Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
            Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,
            Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
            Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
            Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
            Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
            Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
            Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
            Lorensuhera, L., Loulised, H., Lozado, R., Lu, X., Ma, J.,
            Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
            Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
            Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
            Milobavjevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
            Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
            Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
            Nwaokemeluh, O., Okwuonon, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
            Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
            Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
            Fuzao, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
            Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
            Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,

```

```

Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, S., Wang, D.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlaczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 204512)
Worley, K.C.
Direct Submission
Submitted (24-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 204512)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23908182.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVZK
Center clone name: CH230-50113
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 16682 bases at least Q40
Consensus quality: 170683 bases at least Q30
Consensus quality: 173177 bases at least Q20
Estimated insert size: 174509; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 204512: contig of 204512 bp in length.
Location/Qualifiers
1..204512
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-50113"
FEATURES
    source

```

```

misc_feature 1. .1705
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 3869. .4734
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:B2277134"
64608. .66128
/note="wgs_contig"
misc_feature 166948. .168569
/note="wgs_contig"
misc_feature 168900. .170461
/note="wgs_contig"
misc_feature 192738. .193528
/note="clone_boundary
clone_end:T7
site:
end_sequence:B2277133"
196728. .198414
/note="wgs_end_extension
clone_end:T7"
misc_feature 199983. .201685
/note="wgs_end_extension
clone_end:T7"
misc_feature 201736. .202797
/note="wgs_end_extension
clone_end:T7"
misc_feature 203408. .204512
/note="wgs_end_extension
clone_end:T7"

ORIGIN
Query Match 80.8%; Score 20.2; DB 2; Length 204512;
Best Local Similarity 88.0%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAATTCAGTTCATAGCTTGT 25
|||||
DB 198373 TGAGAAATTCATTTTCATAGTGT 198349

RESULT 11
BX927073 215185 bp DNA linear HTG 18-FEB-2004
LOCUS Danio rerio clone DKEYP-78C2, *** SEQUENCING IN PROGRESS ***, 7
DEFINITION unordered pieces.
ACCESSION BX927073.3 GI:42592534
VERSION BX927073.3
KEYWORDS HTG; HTGS PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 215185)
McLay, K.
Direct Submission
Submitted (16-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 17, 2004 this sequence version replaced gi:41016204.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zKp78C2
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 211838 bases at least Q40

misc_feature 1. .1705
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 3869. .4734
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:B2277134"
64608. .66128
/note="wgs_contig"
misc_feature 166948. .168569
/note="wgs_contig"
misc_feature 168900. .170461
/note="wgs_contig"
misc_feature 192738. .193528
/note="clone_boundary
clone_end:T7
site:
end_sequence:B2277133"
196728. .198414
/note="wgs_end_extension
clone_end:T7"
misc_feature 199983. .201685
/note="wgs_end_extension
clone_end:T7"
misc_feature 201736. .202797
/note="wgs_end_extension
clone_end:T7"
misc_feature 203408. .204512
/note="wgs_end_extension
clone_end:T7"

FEATURES
source 1. .215185
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEYP-78C2"
/clone_lib="DanioKeyPilot"
1. .3729
/note="assembly_fragment:00023
fragment chain:1"
misc_feature 3830. .26383
/note="assembly_fragment:00331
fragment chain:1"
misc_feature 26484. .33119
/note="assembly_fragment:00064
fragment chain:2"
misc_feature 33220. .110818
/note="assembly_fragment:01713
fragment chain:2"
misc_feature 110319. .127149
/note="assembly_fragment:00146"
127250. .154645
/note="assembly_fragment:00622.0"
misc_feature 154746. .215185
/note="assembly_fragment:00949"

ORIGIN
Query Match 80.8%; Score 20.2; DB 2; Length 215185;
Best Local Similarity 88.0%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAATTCAGTTCATAGCTTGT 25
|||||
DB 124798 TGAGAAATTCAGTTTATAGCTTTT 124822

RESULT 12
AC098027
LOCUS Rattus norvegicus clone CH230-53E3, *** SEQUENCING IN PROGRESS ***,
DEFINITION 2 unordered pieces.
ACCESSION AC098027
VERSION AC098027.6 GI:30521644
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)

```

Consensus quality: 212487 bases at least Q30
 Consensus quality: 213129 bases at least Q20
 Insert size: 214585; sum-of-contigs
 Insert size: 219324; 2.6% error; agarose-fp
 Quality coverage: 8.00x in Q20 bases; sum-of-contigs Quality
 coverage: 7.92x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

1 3729: contig of 3729 bp in length
 * 3730 3829: gap of 100 bp
 * 3830 26383: contig of 22554 bp in length
 * 26384 26483: gap of 100 bp
 * 26484 33119: contig of 6636 bp in length
 * 33120 33219: gap of 100 bp
 * 33220 110818: contig of 77599 bp in length
 * 110819 110918: gap of 100 bp
 * 110919 127149: contig of 16231 bp in length
 * 127150 127249: gap of 100 bp
 * 127250 154645: contig of 27396 bp in length
 * 154646 154745: gap of 100 bp
 * 154746 215185: contig of 60440 bp in length.

Location/Qualifiers
 1. .215185
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEYP-78C2"
 /clone_lib="DanioKeyPilot"

misc_feature 1. .3729
 /note="assembly_fragment:00023
 fragment chain:1"

misc_feature 3830. .26383
 /note="assembly_fragment:00331
 fragment chain:1"

misc_feature 26484. .33119
 /note="assembly_fragment:00064
 fragment chain:2"

misc_feature 33220. .110818
 /note="assembly_fragment:01713
 fragment chain:2"

misc_feature 110319. .127149
 /note="assembly_fragment:00146"

misc_feature 127250. .154645
 /note="assembly_fragment:00622.0"

misc_feature 154746. .215185
 /note="assembly_fragment:00949"

ORIGIN
 Query Match 80.8%; Score 20.2; DB 2; Length 215185;
 Best Local Similarity 88.0%; Pred. No. 1.9e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAATTCAGTTCATAGCTTGT 25
 |||||

DB 124798 TGAGAAATTCAGTTTATAGCTTTT 124822

RESULT 12

AC098027
 LOCUS Rattus norvegicus clone CH230-53E3, *** SEQUENCING IN PROGRESS ***,
 DEFINITION 2 unordered pieces.

ACCESSION AC098027

VERSION AC098027.6 GI:30521644

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 226228)
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Blewalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gbregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuewa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokeleleh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, P., Polindexter, A., Popovic, D., Primus, E., Pu, L. L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savaryshev, A., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Soes, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G., and Gibbs, R. A.

TITLE

Unpublished
Direct Submission

REFERENCE
AUTHORS

2 (bases 1 to 226228)
Worley, K. C.

TITLE

Direct Submission

JOURNAL

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

3 (bases 1 to 226228)
Rat Genome Sequencing Consortium.

TITLE

Direct Submission

JOURNAL

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 10, 2003 this sequence version replaced gi:23096288.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GFYN

Center clone name: CH230-53E3

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 217218 bases at least Q40

Consensus quality: 219723 bases at least Q30

Consensus quality: 221029 bases at least Q20

Estimated insert size: 228733; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved

* 1 224885: contig of 224885 bp in length

* 224886 224985: gap of unknown length

* 224986 226228: contig of 1243 bp in length.

----- Location/Qualifiers

source

1. 226228

/organism="Rattus norvegicus"

/mol_type="Genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-53E3"

1. 1471

/note="wgs contig"

4273. 6861

/note="wgs_contig"

ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 226228;
Best Local Similarity 88.0%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCTGT 25

|||||

Db 183838 TGAAGAAATTCATTCATAGAGTGT 183862

RESULT 13

AR438843/c

LOCUS

DEFINITION

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843

AR438843


```

FEATURES                                     Location/Qualifiers
Source                                     1..44848
/mol_type="genomic DNA"

ORIGIN
Query Match                               79.2%; Score 19.8; DB 6; Length 44848;
Best Local Similarity 91.3%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24
|||||
Db 30803 GAAGAAATTCAGTTCATGTTG 30781

RESULT 14
AC093382                                     57633 bp DNA linear PRI 10-JAN-2002
LOCUS Homo sapiens BAC clone RP11-476D14 from 2, complete sequence.
DEFINITION AC093382
ACCESSION AC093382
VERSION AC093382.3 GI:17978444
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE
AUTHORS Radonienko,M., Kozlowicz,A. and Doebber,A.
TITLE The sequence of Homo sapiens BAC clone RP11-476D14
JOURNAL Unpublished (2002)
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (26-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Dec 26, 2001 this sequence version replaced gi:17227294.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0476D14
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-504L12, 2000 bp overlap finished in RP11-476D14. Actual start of this clone is at base position 1 of RP11-476D14; actual end is at base position 106332 of RP11-504L12.

FEATURES Location/Qualifiers

```

1..57633
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-476D14"
/clone_lib="RPCI-11"
217..580
/rpt_family="L2"
581..879
/rpt_family="Alu"
880..1165
/rpt_family="L2"
1279..1413
/rpt_family="MIR"
1572..2071
/rpt_family="MaLR"
2133..2163
/rpt_family="(TTTC)n"
2137..2447
/rpt_family="Alu"
3163..3423
/rpt_family="Alu"
3418..3450
/rpt_family="AT-rich"
3437..3712
/rpt_family="L1"
3547..3588
/rpt_family="(TG)n"
3923..3945
/rpt_family="AT-rich"
4058..4291
/rpt_family="MER1_type"
4309..4553
/rpt_family="L1"
4575..5148
/rpt_family="L1"
5136..5173
/rpt_family="AT-rich"
6280..6376
/rpt_family="MER1_type"

```



```

repeat_region 6789. .6921
/rpt_family="MIR"
repeat_region 6922. .7276
/rpt_family="MaLR"
repeat_region 8012. .8033
/rpt_family="AT_rich"
repeat_region 8015. .8308
/rpt_family="Alu"
repeat_region 8900. .9227
/rpt_family="L2"
repeat_region 11593. .11747
/rpt_family="CR1"
repeat_region 13142. .13247
/rpt_family="MIR"
repeat_region 13782. .13898
/rpt_family="MIR"
repeat_region 15262. .15515
/rpt_family="MIR"
repeat_region 15577. .15642
/rpt_family="MER1_type?"
repeat_region 15647. .15837
/rpt_family="MIR"
misc_feature 15974. .16439
/notes="similar to Mus musculus EST BB614050
(NID:G16454511)"
repeat_region 17442. .18075
/rpt_family="ERV1"
repeat_region 19208. .19494
/rpt_family="ERV1"
repeat_region 19495. .20184
/rpt_family="ERV1"
repeat_region 20185. .20438
/rpt_family="ERV1"
repeat_region 20449. .20592
/rpt_family="L1"
repeat_region 20932. .21491
/rpt_family="ERV1"
repeat_region 21492. .21533
/rpt_family="AT_rich"
repeat_region 23942. .24226
/rpt_family="Alu"
repeat_region 24369. .24735
/rpt_family="ERV1"
repeat_region 25188. .25292
/rpt_family="MaLR"
repeat_region 25293. .25606
/rpt_family="Alu"
repeat_region 25450. .25481
/rpt_family="(T)n"
misc_feature 25457. .25472
/notes="match to EST BGI67210 (NID:G12673913)"
repeat_region 25607. .25924
/rpt_family="MaLR"
repeat_region 26111. .26503
/rpt_family="ERV1"
repeat_region 26667. .27022
/rpt_family="ERV1"
repeat_region 27061. .27121
/rpt_family="ERV1"
repeat_region 27131. .27355
/rpt_family="ERV1"
repeat_region 27651. .27720
/rpt_family="Alu"
repeat_region 27795. .27847
/rpt_family="L2"
repeat_region 28053. .28276
/rpt_family="Alu"
repeat_region 29103. .29243
/rpt_family="(GAAA)n"
repeat_region 29284. .29501
/rpt_family="MER1_type"

```

Query Match

79.2%; Score 19.8; DB 9; Length 57633;

```

Best Local Similarity 91.3%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTTCATGCTTG 24
|||||
Db 19627 GAAGAAATTCAGTTTCATGCTTG 19649

RESULT 15
AC016347 72040 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens chromosome 15 clone RP11-189L10 map 15, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
AC016347
AC016347.2 GI:9141889
VERSION HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 72040)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 15, clone RP11-189L10
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 72040)
REFERENCE Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Castle, A., Collangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Garg, J.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karats, A., Klein, J.,
Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6466986.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4952
Center clone name: 189_L_10
-----
* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 807: contig of 807 bp in length
* 808 907: gap of 100 bp
* 908 1700: contig of 793 bp in length
* 1701 1800: gap of 100 bp
* 1801 2587: contig of 787 bp in length
* 2588 2687: gap of 100 bp
* 2688 3476: contig of 789 bp in length

```

```
* 3477 3576: gap of 100 bp
* 3577 4312: contig of 736 bp in length
* 4313 4412: gap of 100 bp
* 4413 5209: contig of 797 bp in length
* 5210 5309: gap of 100 bp
* 5310 6079: contig of 770 bp in length
* 6080 6179: gap of 100 bp
* 6180 6971: contig of 792 bp in length
* 6972 7071: gap of 100 bp
* 7072 7875: contig of 804 bp in length
* 7876 7975: gap of 100 bp
* 7977 8752: contig of 777 bp in length
* 8753 8852: gap of 100 bp
* 8853 9634: contig of 782 bp in length
* 9635 9734: gap of 100 bp
* 9735 10519: contig of 785 bp in length
* 10520 10619: gap of 100 bp
* 10620 11415: contig of 796 bp in length
* 11416 11515: gap of 100 bp
* 11516 12314: contig of 799 bp in length
* 12315 12414: gap of 100 bp
* 12415 13201: contig of 787 bp in length
* 13202 13301: gap of 100 bp
* 13302 14092: contig of 791 bp in length
* 14093 14192: gap of 100 bp
* 14193 14981: contig of 789 bp in length
* 14982 15081: gap of 100 bp
* 15082 15849: contig of 768 bp in length
* 15850 15949: gap of 100 bp
* 15950 16775: contig of 826 bp in length
* 16776 16875: gap of 100 bp
* 16876 17667: contig of 792 bp in length
* 17668 17767: gap of 100 bp
* 17768 18574: contig of 807 bp in length
* 18575 18674: gap of 100 bp
* 18675 19478: contig of 804 bp in length
* 19479 19578: gap of 100 bp
* 19579 20380: contig of 802 bp in length
* 20381 20480: gap of 100 bp
* 20481 21287: contig of 807 bp in length
* 21288 21387: gap of 100 bp
* 21388 22175: contig of 788 bp in length
* 22176 22275: gap of 100 bp
* 22276 23029: contig of 754 bp in length
* 23030 23129: gap of 100 bp
* 23130 23914: contig of 785 bp in length
* 23915 24014: gap of 100 bp
* 24015 24788: contig of 774 bp in length
* 24789 24888: gap of 100 bp
* 24889 25686: contig of 798 bp in length
* 25687 25786: gap of 100 bp
* 25787 26593: contig of 807 bp in length
* 26594 27521: contig of 828 bp in length
* 27522 27621: gap of 100 bp
* 27622 28388: contig of 767 bp in length
* 28389 28488: gap of 100 bp
* 28489 29298: contig of 810 bp in length
* 29299 29398: gap of 100 bp
* 29399 30193: contig of 795 bp in length
* 30194 30293: gap of 100 bp
* 30294 31059: contig of 776 bp in length
* 31070 31169: gap of 100 bp
* 31170 31969: contig of 800 bp in length
* 31970 32069: gap of 100 bp
* 32070 32847: contig of 778 bp in length
* 32848 32947: gap of 100 bp
* 32948 33738: contig of 791 bp in length
* 33739 33838: gap of 100 bp
* 33839 34668: contig of 830 bp in length
* 34669 34768: gap of 100 bp
* 34769 35568: contig of 800 bp in length
* 35569 35668: gap of 100 bp
```

```
* 35669 36472: contig of 804 bp in length
* 36473 36572: gap of 100 bp
* 36573 37360: contig of 788 bp in length
* 37361 37460: gap of 100 bp
* 37461 38267: contig of 807 bp in length
* 38268 38367: gap of 100 bp
* 38368 39157: contig of 790 bp in length
* 39158 39257: gap of 100 bp
* 39258 40034: contig of 777 bp in length
* 40035 40134: gap of 100 bp
* 40135 40933: contig of 799 bp in length
* 40934 41033: gap of 100 bp
* 41034 41825: contig of 792 bp in length
* 41826 41925: gap of 100 bp
* 41926 42701: contig of 776 bp in length
* 42702 42801: gap of 100 bp
* 42802 43589: contig of 788 bp in length
* 43590 43689: gap of 100 bp
* 43690 44499: contig of 810 bp in length
* 44500 44599: gap of 100 bp
* 44600 45389: contig of 790 bp in length
* 45390 45489: gap of 100 bp
* 45490 46284: contig of 795 bp in length
* 46285 46384: gap of 100 bp
* 46385 47191: contig of 807 bp in length
* 47192 47291: gap of 100 bp
* 47292 48060: contig of 789 bp in length
* 48061 48160: gap of 100 bp
* 48161 48954: contig of 794 bp in length
* 48955 49054: gap of 100 bp
* 49055 49838: contig of 784 bp in length
* 49839 49938: gap of 100 bp
* 49939 50727: contig of 789 bp in length
* 50728 50827: gap of 100 bp
* 50828 51634: contig of 807 bp in length
* 51635 51734: gap of 100 bp
* 51735 52530: contig of 796 bp in length
* 52531 52630: gap of 100 bp
* 52631 53412: contig of 782 bp in length
* 53413 53512: gap of 100 bp
* 53513 54297: contig of 785 bp in length
* 54298 54397: gap of 100 bp
* 54398 55184: contig of 787 bp in length
* 55185 55284: gap of 100 bp
* 55285 56093: contig of 809 bp in length
* 56094 56193: gap of 100 bp
* 56194 57001: contig of 808 bp in length
* 57002 57101: gap of 100 bp
* 57102 57900: contig of 799 bp in length
* 57901 58000: gap of 100 bp
* 58001 58747: contig of 747 bp in length
* 58748 58847: gap of 100 bp
* 58848 59641: contig of 794 bp in length
* 59642 59741: gap of 100 bp
* 59742 60536: contig of 795 bp in length
* 60537 60636: gap of 100 bp
* 60637 61411: contig of 775 bp in length
* 61412 61511: gap of 100 bp
* 61512 62308: contig of 797 bp in length
* 62309 62408: gap of 100 bp
* 62409 63215: contig of 807 bp in length
* 63216 63315: gap of 100 bp
* 63316 64099: contig of 784 bp in length
* 64100 64199: gap of 100 bp
```

Query Match 79.2%; Score 19.8; DB 2; Length 72040;
Beat Local Similarity 91.3%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24
|||||
Db 71709 GAAGAAATTCAGTTCATAGCTTG 71731
|||||

Search completed: August 13, 2005, 05:04:24
Job time : 797.298 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 12, 2005, 22:21:57 ; Search time 200.266 Seconds
(without alignments)
738.985 Million cell updates/sec

Title: US-10-673-854-3
Perfect score: 25
Sequence: 1 tgaagaattcagttcatagctgtg 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| C 1 | 19.8 | 79.2 | 922 | 12 | ADJ43067 |
| C 2 | 19.8 | 79.2 | 44848 | 3 | Aaa75080 Nucleotid |
| C 3 | 19.8 | 79.2 | 44848 | 10 | Adg88832 Human bpa |
| C 4 | 19.8 | 79.2 | 44848 | 12 | Adl16411 Human hep |
| C 5 | 19.8 | 79.2 | 44848 | 12 | Adm48748 Human bpa |
| C 6 | 19.8 | 79.2 | 183178 | 10 | Adl13873 Osteoarth |
| C 7 | 19.2 | 76.8 | 49507 | 11 | Acn44572 Mouse gen |
| C 8 | 19.2 | 76.8 | 202001 | 6 | Abss2506 Human tra |
| C 9 | 19.2 | 76.8 | 202001 | 10 | Adg46742 Human tra |
| C 10 | 19 | 76.0 | 235 | 12 | Adp92027 Cotton ex |
| C 11 | 18.8 | 75.2 | 1227 | 11 | Ach97547 Klebsiell |
| C 12 | 18.8 | 75.2 | 176771 | 12 | Adg97167 Human can |
| C 13 | 18.6 | 74.4 | 4212 | 6 | Abz14865 Arabidops |
| C 14 | 18.6 | 74.4 | 116858 | 11 | Acn44212 Mouse gen |
| C 15 | 18.2 | 72.8 | 978 | 10 | Adk54409 Plant DNA |
| C 16 | 18.2 | 72.8 | 5536 | 4 | Aak89473 Human dig |
| C 17 | 18.2 | 72.8 | 40463 | 11 | Acn44904 Mouse gen |
| C 18 | 18.2 | 72.8 | 98638 | 12 | Adg97919 Mouse can |
| C 19 | 18.2 | 72.8 | 164991 | 10 | Adl13635 Osteoarth |
| C 20 | 18.2 | 72.8 | 176080 | 12 | Adl08124 Human gen |

| | | | | | |
|------|------|------|--------|----|--------------------|
| C 21 | 18 | 72.0 | 520 | 5 | AAS84456 |
| C 22 | 18 | 72.0 | 2509 | 12 | ADQ63965 |
| C 23 | 17.8 | 71.2 | 491 | 4 | ABA57693 Human foe |
| C 24 | 17.8 | 71.2 | 491 | 4 | AAI37269 |
| C 25 | 17.8 | 71.2 | 491 | 4 | AAK31375 Human bon |
| C 26 | 17.8 | 71.2 | 491 | 4 | AAK05751 Human bra |
| C 27 | 17.8 | 71.2 | 491 | 4 | ABS31056 Human liv |
| C 28 | 17.8 | 71.2 | 491 | 6 | ABS06128 Human gen |
| C 29 | 17.8 | 71.2 | 708 | 13 | ADS60583 |
| C 30 | 17.8 | 71.2 | 1325 | 4 | AAH52803 |
| C 31 | 17.8 | 71.2 | 1485 | 6 | ABN91876 |
| C 32 | 17.8 | 71.2 | 1485 | 13 | ADS01312 |
| C 33 | 17.8 | 71.2 | 2978 | 4 | AAH54735 |
| C 34 | 17.8 | 71.2 | 3745 | 4 | AAH54681 S. epider |
| C 35 | 17.6 | 70.4 | 384 | 8 | ABX50480 Bovine ES |
| C 36 | 17.6 | 70.4 | 462 | 6 | ABN26076 Human ORF |
| C 37 | 17.6 | 70.4 | 515 | 12 | ADJ10653 Recombina |
| C 38 | 17.6 | 70.4 | 545 | 13 | ADQ50923 Novel can |
| C 39 | 17.6 | 70.4 | 1259 | 6 | ABA01916 Human cel |
| C 40 | 17.6 | 70.4 | 1999 | 10 | ADC08531 Rice DNA |
| C 41 | 17.6 | 70.4 | 3089 | 13 | ADS54895 Bacterial |
| C 42 | 17.6 | 70.4 | 3270 | 13 | ACN38754 Tumour-as |
| C 43 | 17.6 | 70.4 | 3808 | 5 | ADL62420 Human ova |
| C 44 | 17.6 | 70.4 | 90220 | 6 | ABK83576 Human CDN |
| C 45 | 17.6 | 70.4 | 245531 | 13 | ABD33022 Human can |

ALIGNMENTS

RESULT 1

ADJ43067/c
ID ADJ43067 standard; CDNA; 922 BP.

AC ADJ43067;

DT 06-MAY-2004 (first entry)

DE Plant cDNA #4067.

KW Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW antifungal.

OS Eukaryota.

XX XX

PN US2004016025-A1.

XX XX

PD 22-JAN-2004.

XX XX

PF 26-SEP-2002; 2002US-00260238.

XX XX

PR 26-SEP-2001; 2001US-0325277P.

PR 26-SEP-2001; 2001US-0325448P.

PR 04-APR-2002; 2002US-0370620P.

XX XX

PA (BUDW/) BUDWORTH P.

PA (MOUG/) MOUGHAMER T.

PA (BRIG/) BRIGGS S P.

PA (COOP/) COOPER B.

PA (GLAZ/) GLAZEBROOK J.

PA (GOFF/) GOFF S A.

PA (KATA/) KATAGIRI F.

PA (KREP/) KREPS J.

PA (PROV/) PROVART N.

PA (RICK/) RICHE D.

PA (ZHUT/) ZHU T.

XX XX

PI Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;

PI Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;

```
XX WPI; 2004-190374/18.
XX
XX New rice promoter, useful for manipulating crop plants to alter or
PT improve phenotypic characteristics, e.g. produce large quantities of oil
PT or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT or high nutritional value.
XX
XX Example 13; SEQ ID NO 4067; 230pp; English.
XX
XX The invention relates to plant nucleotide sequences that direct seed-,
CC leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC or constitutive transcription of an operatively linked nucleic acid
CC segment. The invention also relates to a method for augmenting a plant
CC genome and a method of identifying a gene, where its expression is
CC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC encode are useful for manipulating crop plants to alter or improve
CC phenotypic characteristics, to produce large quantities of oil or
CC proteins, to incur resistance to insecticides, viruses or fungi, and to
CC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC have a high nutritional value with reduced apical dominance or dwarfism,
CC early flowering or altered metabolic pathways. This sequence represents a
CC plant nucleic acid of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification but was obtained in
CC electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 922 BP; 210 A; 294 C; 234 G; 184 T; 0 U; 0 Other;
SQ
Query Match 79.2%; Score 19.8; DB 12; Length 922;
Best Local Similarity 91.3%; Pred. No. 64;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAAGAAATTCAGTTTCATAGCTTG 24
Db 869 GAAGAAATTCAGTTCTCAGCATG 847
RESULT 2
AAA75080/C
ID AAA75080 standard; DNA; 44848 BP.
XX
AC AAA75080;
XX
XX 15-JAN-2001 (first entry)
XX
XX Nucleotide sequence of the human heparanase gene.
XX
XX Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.
XX
XX Homo sapiens.
XX
XX WO200052178-A1.
XX
XX 08-SEP-2000.
XX
XX 14-FEB-2000; 2000WO-US003542.
XX
XX 01-MAR-1999; 99US-00258892.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodavsky I, Feinstein E;
XX WPI; 2000-579289/54.
XX
```

```
XX New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumor, inflammation, autoimmunity, neurodegenerative diseases.
XX
XX Claim 9; Page 131-143; 152pp; English.
XX
XX The present sequence represents a human gene which encodes a protein with
CC heparanase catalytic activity. The heparanase (hpa) polynucleotide is
CC useful in gene therapy, particularly in treating tumour, inflammation or
CC autoimmunity. Particularly, the polynucleotide is useful in modulating
CC the bioavailability of heparin-binding growth factors, cellular responses
CC to heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.
CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular
CC susceptibility to certain viral and some bacterial and protozoa
CC infections, or disintegration of neurodegenerative plaques. The
CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or
CC radiation burns), and in the treatment of angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC bacterial or protozoa infections
XX
XX Sequence 44848 BP; 12560 A; 9646 C; 8930 G; 13712 T; 0 U; 0 Other;
SQ
Query Match 79.2%; Score 19.8; DB 3; Length 44848;
Best Local Similarity 91.3%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAAGAAATTCAGTTTCATAGCTTG 24
Db 30803 GAAGAAATTCAGTTTCATGTTTG 30781
RESULT 3
ADG88832/C
ID ADG88832 standard; DNA; 44848 BP.
XX
AC ADG88832;
XX
XX 11-MAR-2004 (first entry)
XX
XX Human hpa genomic DNA.
XX
XX Wound healing; heparanase; ulcer; burn; laceration; surgical incision;
KW necrosis; pressure wound; diabetic ulcer; angiogenesis; human; therapy;
KW gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 2743..41863
FT /*tag= a
FT /product= "Human hpa protein"
XX
XX US2003161823-A1.
XX
XX 28-AUG-2003.
XX
XX 14-JAN-2003; 2003US-00341582.
XX
XX 31-AUG-1998; 98WO-US017954.
XX
XX 01-MAR-1999; 99US-00258892.
XX
XX 06-FEB-2001; 2001US-00776874.
XX
XX 05-SEP-2001; 2001WO-IL000830.
XX
XX 19-NOV-2001; 2001US-00988113.
XX
XX (ILAN/) ILAN N.
PA (VLOD/) VLODAVSKY I.
PA (YACO/) YACOBY-ZEEVI O.
PA (PECK/) PECKER I.
PA (FEIN/) FEINSTEIN E.
XX
XX Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;
```

```

XX WPI; 2003-897910/82.
DR P-PSDB; ADG88800.
XX
XX Composition for treating a wound comprising recombinant heparanase is
PT useful to induce or accelerate wound healing and induce or accelerate
PT angiogenesis.
XX
XX Claim 4; SEQ ID NO 42; 143pp; English.
XX
XX The present invention relates to methods and compositions for inducing
CC and/or accelerating wound healing via the catalytic activity of
CC heparanase. The invention is used to induce or accelerate a healing
CC process, particularly of an ulcer, burn, laceration, surgical incision,
CC necrosis, pressure wound, diabetic ulcer and to induce or accelerate
CC angiogenesis. The present sequence is human hpa genomic DNA.
XX
XX Sequence 44848 BP; 12560 A; 9646 C; 8930 G; 13712 T; 0 U; 0 Other;
SQ
Query Match 79.2%; Score 19.8; DB 10; Length 44848;
Best Local Similarity 91.3%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAAGAAATTCAGTTCATAGCTTG 24
DB 30803 GAAGAAATTCAGTTCATAGCTTG 30781
RESULT 4
ADL16411/C
ID ADL16411 standard; DNA; 44848 BP.
XX
XX ADL16411;
XX
XX 06-MAY-2004 (first entry)
XX
XX Human heparanase genomic DNA.
DE
XX
XX Human; ds; heparanase; gene; heparanase-dependent cancer; cancer;
XX autoimmune reaction; inflammation; chromosome 4.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX promoter 2535..2635
XX /*tag= a
XX /note= "Minimal promoter region. Claimed in claim 15"
XX 5'UTR 2635..2742
XX /*tag= b
XX /note= "Claimed in claim 19"
XX 3'UTR 41864..41890
XX /*tag= c
XX /note= "Claimed in claim 20"
XX
XX US2003236215-A1.
XX
XX 25-DEC-2003.
XX
XX 09-JUN-2003; 2003US-00456573.
XX
XX 31-AUG-1998; 98WO-US017954.
XX 01-MAR-1999; 99US-00258892.
XX 08-NOV-1999; 99US-00435739.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
XX Pecker I, Vlodavsky I, Feinstein B;
XX
XX WPI; 2004-070610/07.
XX
XX New antisense oligonucleotide hybridizable with a polynucleotide encoding
PT a polypeptide with heparanase activity, useful for treating diseases such

```

```

PT as cancer and autoimmune disorders.
XX
XX Claim 2; SEQ ID NO 42; 108pp; English.
XX
XX The invention relates to an antisense oligonucleotide (ASO) comprising a
CC polynucleotide or a polynucleotide analogue of at least 10 bases being
CC hybridizable in vivo, under physiological conditions, with a portion of
CC a polynucleotide strand encoding a polypeptide having heparanase
CC catalytic activity. Also included are a method of in vivo downregulating
CC heparanase activity (comprising administering the ASO in vivo), a method
CC of treating a subject suffering from a pathological condition
CC (characterised by heparanase activity, comprising administering ASO to
CC the subject), a pharmaceutical composition comprising the ASO and a
CC carrier, an antisense nucleic acid construct (comprising a promoter
CC sequence and a polynucleotide sequence directing the synthesis of an
CC antisense RNA sequence of at least 10 bases being hybridisable in vivo,
CC under physiological conditions, with a polynucleotide strand encoding a
CC polypeptide having heparanase catalytic activity), a method of in vivo
CC downregulating heparanase activity (comprising administering in vivo the
CC antisense nucleic acid construct), a pharmaceutical composition
CC comprising the antisense nucleic acid construct and a carrier, and an
CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide
CC analogue of at least 10 bases being hybridisable in vivo, under
CC physiological conditions, with a portion of a polynucleotide strand being
CC characterised by forming at least a portion of an untranslated region
CC (UTR) for a polynucleotide strand encoding a polypeptide having
CC heparanase catalytic activity. The methods and compositions of the
CC present invention are useful for the prevention and/or treatment of
CC diseases or conditions associated with aberrant heparanase activity, such
CC as heparanase-dependent cancer, cancer, autoimmune reaction and
CC inflammation. The gene for human heparanase is located on chromosome 4.
XX
XX The present sequence is the human heparanase gene.
XX
XX Sequence 44848 BP; 12560 A; 9646 C; 8930 G; 13712 T; 0 U; 0 Other;
SQ
Query Match 79.2%; Score 19.8; DB 12; Length 44848;
Best Local Similarity 91.3%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAAGAAATTCAGTTCATAGCTTG 24
DB 30803 GAAGAAATTCAGTTCATAGCTTG 30781
RESULT 5
ADM48748/C
ID ADM48748 standard; DNA; 44848 BP.
XX
XX ADM48748;
XX
XX 03-JUN-2004 (first entry)
XX
XX Human hpa genomic DNA.
XX
XX Transgenic animal; heparanase; cancer; viral infection; restenosis;
XX neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
XX human; gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX exon 2635..2969
XX /*tag= a
XX /number= 1
XX CDS 2743..41863
XX /*tag= b
XX /product= "Hpa protein"
XX intron 2970..15141
XX /*tag= c
XX /number= 1
XX exon 15142..15287
XX /*tag= d
XX /number= 2

```

```
FT intron 15288..18035
FT      /*tag= e
FT      /number= 2
FT exon 18036..18161
FT      /*tag= f
FT intron 18162..24214
FT      /*tag= g
FT      /number= 3
FT exon 24215..24388
FT      /*tag= h
FT intron 24389..26600
FT      /*tag= i
FT exon 26601..26769
FT      /*tag= j
FT intron 26770..27413
FT      /*tag= k
FT exon 27414..27461
FT      /*tag= l
FT intron 27462..27996
FT      /*tag= m
FT      /number= 6
FT exon 27997..28090
FT      /*tag= n
FT intron 28091..28540
FT      /*tag= o
FT      /number= 7
FT exon 28541..28647
FT      /*tag= p
FT intron 28648..31173
FT      /*tag= q
FT exon 31174..31288
FT      /*tag= r
FT intron 31289..35227
FT      /*tag= s
FT      /number= 9
FT exon 35228..35346
FT      /*tag= t
FT intron 35347..36389
FT      /*tag= u
FT exon 36390..36536
FT      /*tag= v
FT      /number= 11
FT intron 36537..41703
FT      /*tag= w
FT exon 41704..41890
FT      /*tag= x
FT      /number= 12
XX US2003217375-A1.
XX
XX PD 20-NOV-2003.
XX
XX PF 24-FEB-2003; 2003US-00371218.
XX
XX PR 31-AUG-1998; 98WO-US017954.
XX PR 01-MAR-1999; 99US-00258892.
XX PR 06-FEB-2001; 2001US-00776874.
XX PR 19-NOV-2001; 2001US-00988113.
XX
XX PA (ZCHAJ/) ZCHARIA E.
PA (VLOD/) VLODAVSKY I.
PA (METZ/) METZGER S.
PA (PECK/) PECKER I.
PA (ILAN/) ILAN N.
PA (CHAJ/) CHAJEK-SHAUL T.
PA (GOLD/) GOLDSHMIDT O.
XX
PI Zcharia E, Vlodevsky I, Metzger S, Pecker I, ilan N;
PI Chajek-Shaul T, Goldshmidt O;
XX WPI; 2004-021918/02.
XX P-PSDB; ADM48759.
XX
PT New transgenic non-human animal expressing heparinase, useful as models
PT for human disease, such as cancers, viral infection, neurodegenerative
PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
PS Example 10; SEQ ID NO 42; 106pp; English.
XX
CC The present invention relates to a transgenic non-human animal whose
CC genome comprises an exogenous polynucleotide sequence, including a
CC promoter active in tissues of the non-human, a region encoding a human
CC heparanase, where the promoter and the region encoding human heparanase
CC are operably linked in the exogenous polynucleotide such that human
CC heparanase is expressed in at least a portion of the cells of the non-
CC human animal. The methods and compositions of the present invention are
CC useful for the production of transgenic animals expressing heparanase, to
CC be used as models for human diseases such as cancers, viral infection,
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
CC disorders. The present sequence is human hpa genomic DNA used in the
CC exemplification of the invention.
XX
SQ Sequence 44848 BP; 12560 A; 9646 C; 8930 G; 13712 T; 0 U; 0 Other;
Query Match 79.2%; Score 19.8; DB 12; Length 44848;
Best Local Similarity 91.3%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAAGAAATTCAGTTCATAGCTTG 24
Db 30803 GAAGAAATTCAGTTCATAGCTTG 30781
RESULT 6
ADL13873/c
ID ADL13873 standard; DNA; 183178 BP.
XX
AC ADL13873;
XX
XX 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #405.
XX
KW ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW joint space narrowing; osteophyte development; joint pain;
KW osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
XX WO2003054166-A2.
XX
XX 03-JUL-2003.
XX
XX PF 19-DEC-2002; 2002WO-US041225.
XX
XX PR 20-DEC-2001; 2001US-0342603P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Jones KA, Schafer A;
XX WPI; 2003-559141/52.
XX
```


PT Determining susceptibility of an individual to joint space narrowing,
 PT osteophyte development and/or joint pain comprises identifying whether
 PT the individual has at least one polymorphism in a polynucleotide encoding
 PT a protein.

XX Disclosure; SEQ ID NO 405; 297pp; English.

CC The invention relates to a method of determining susceptibility of an
 CC individual to joint space narrowing and/or osteophyte development and/or
 CC joint pain comprising identifying whether the individual has at least one
 CC polymorphism in a polynucleotide encoding at least one of the protein
 CC listed in the specification. The methods, composition and agent are
 CC useful for modulating the susceptibility of an individual to joint space
 CC narrowing and/or osteophyte development and/or joint pain that is
 CC associated with a disease, preferably osteoarthritis. The cell line and
 CC the non-human animal are useful for screening for an agent for diagnosing
 CC an individual having susceptibility to joint space narrowing and/or
 CC osteophyte development and/or joint pain. This sequence corresponds to
 CC the polynucleotide encoding a protein listed in the specification. (Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences).

SQ Sequence 183178 BP; 52771 A; 36295 C; 36571 G; 54082 T; 0 U; 3459 Other;

Query Match 79.2%; Score 19.8; DB 10; Length 183178;

Best Local Similarity 91.3%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24

DB 155952 GAAGAAATTCAGTTCATAGCTTG 155930

RESULT 7

ID ACN44572 standard; DNA; 49507 BP.

AC ACN44572;

DT 18-NOV-2004 (first entry)

XX Mouse genomic sequence MCG21659.

XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.

XX Mus musculus.

XX WO2003073826-A2.

XX 12-SEP-2003.

XX 28-FEB-2003; 2003WO-US0006235.

XX 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-328604/31.

XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 PT comprises a nucleotide sequence.

PS Claim 1; SEQ ID NO 1087; Opp; English.

XX The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to carcinoma Associated protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing

CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CA coding
 CC sequence. Note: this patent is an equivalent to basic patent
 CC US2002182586A1, for which no sequence data was published

XX SQ Sequence 49507 BP; 14131 A; 9178 C; 9693 G; 15785 T; 0 U; 720 Other;

Query Match 76.8%; Score 19.2; DB 11; Length 49507;

Best Local Similarity 87.5%; Pred. No. 1.8e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCTTG 24

DB 1473 TGAAGAAATTCAGTTCATAGCTTG 1496

RESULT 8

ABS52506/C

ID ABS52506 standard; DNA; 202001 BP.

AC ABS52506;

DT 15-NOV-2002 (first entry)

XX Human transporter protein genomic DNA.

XX Human; gene; ds; transporter protein; cell proliferation;
 KW cell differentiation; cell signaling; sodium bicarbonate cotransporter;
 KW transgenic; ligand transport; drug development; SNP;
 KW single nucleotide polymorphism.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT variation replace(2180,C)
 FT /tag= ax

FT /standard_name= "Single nucleotide polymorphism"
 FT exon 3016..3096

FT /tag= a

FT /number= 1

FT 3097..11617

FT /tag= b

FT /number= 1

FT /cons_splice= (5'site:yes,3'site:no)

FT variation replace(4693,C)

FT /tag= ay

FT /standard_name= "Single nucleotide polymorphism"

FT exon 11618..11690

FT /tag= c

FT /number= 2

FT 11691..37943

FT /tag= d

FT /number= 2

FT /cons_splice= (5'site:no,3'site:no)

FT variation replace(13759,T)

FT /tag= az

FT /standard_name= "Single nucleotide polymorphism"

FT variation replace(17580,A)

FT /tag= ba

FT /standard_name= "Single nucleotide polymorphism"

FT variation replace(17701,A)

FT /tag= bc

FT /standard_name= "Single nucleotide polymorphism"

FT variation replace(18151,C)

FT /tag= bd

FT /standard_name= "Single nucleotide polymorphism"

FT variation replace(21076,C)

FT /tag= be

FT /standard_name= "Single nucleotide polymorphism"

```
FT variation replace(22984,C)
FT /*tag= bf
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(25960, .25961,ATT)
FT /*tag= bg
FT variation replace(25961, .25962,TWT)
FT /*tag= bh
FT variation replace(25962,T)
FT /*tag= bi
FT variation replace(25962,T)
FT /*tag= bj
FT variation replace(25962,T)
FT /*tag= bk
FT variation replace(25962,T)
FT /*tag= bl
FT variation replace(25962,T)
FT /*tag= bm
FT variation replace(25962,T)
FT /*tag= bn
FT exon 37944, .38090
FT /*tag= e
FT intron 38091, .76221
FT /*tag= f
FT /number= 3
FT /cons_splice= (5'site:yes,3'site:no)
FT variation replace(40997,C)
FT /*tag= bo
FT variation replace(45222,C)
FT /*tag= bp
FT variation replace(45890,A)
FT /*tag= bq
FT variation replace(49199,R)
FT /*tag= br
FT variation replace(61696,R)
FT /*tag= bs
FT variation replace(63810,C)
FT /*tag= bt
FT variation replace(64061,T)
FT /*tag= bu
FT variation replace(64186,T)
FT /*tag= bv
FT variation replace(65909,A)
FT /*tag= bw
FT variation replace(66361,T)
FT /*tag= bx
FT variation replace(68545,T)
FT /*tag= by
FT variation replace(70223,T)
FT /*tag= bz
FT variation replace(72117,G)
FT /*tag= ca
FT /standard_name= "Single nucleotide polymorphism"
```

```
FT variation replace(74719, .74721,AA)
FT /*tag= cb
FT /standard_name= "Single nucleotide polymorphism"
FT exon 76222, .76360
FT /*tag= g
FT intron 76361, .91402
FT /*tag= h
FT /number= 4
FT /cons_splice= (5'site:yes,3'site:no)
FT variation replace(77206,A)
FT /*tag= cc
FT variation replace(77426,G)
FT /*tag= cd
FT variation replace(78935,C)
FT /*tag= ce
FT variation replace(89179,G)
FT /*tag= cf
FT variation replace(90081,C)
FT /*tag= cg
FT exon 91403, .91563
FT /*tag= i
FT intron 91564, .99311
FT /*tag= j
FT /number= 5
FT /cons_splice= (5'site:yes,3'site:no)
FT variation replace(96033,C)
FT /*tag= ch
FT variation replace(96808,C)
FT /*tag= ci
FT exon 99312, .99500
FT /*tag= k
FT intron 99501, .108867
FT /*tag= l
FT /number= 6
FT /cons_splice= (5'site:yes,3'site:no)
FT variation replace(102300,W)
FT /*tag= cj
FT variation replace(105400,G)
FT /*tag= ck
FT variation replace(105494,C)
FT /*tag= cl
FT variation replace(105911,T)
FT /*tag= cm
FT exon 108868, .108959
FT /*tag= m
FT intron 108960, .110489
FT /*tag= n
FT /number= 7
FT variation replace(110376,C)
FT /*tag= cn
FT exon 110490, .110579
FT /*tag= o
FT intron 110580, .115705
FT /*tag= p
FT variation replace(111250,A)
```



```
FT /standard name= "Single nucleotide polymorphism"
FT replace(77426,G)
FT /*tag= aq
FT /standard name= "Single nucleotide polymorphism"
FT replace(78935,C)
FT /*tag= ar
FT /standard name= "Single nucleotide polymorphism"
FT replace(89179,G)
FT /*tag= as
FT /standard name= "Single nucleotide polymorphism"
FT replace(90081,C)
FT /*tag= at
FT /standard name= "Single nucleotide polymorphism"
FT 91403..91563
FT /*tag= au
FT 91564..99311
FT /*tag= av
FT replace(96033,C)
FT /*tag= aw
FT /standard name= "Single nucleotide polymorphism"
FT replace(96808,C)
FT /*tag= ax
FT 99312..99500
FT /*tag= ay
FT 99501..108867
FT /*tag= az
FT replace(102300,A, T)
FT /*tag= ba
FT /standard name= "Single nucleotide polymorphism"
FT replace(105400,G)
FT /*tag= bb
FT /standard name= "Single nucleotide polymorphism"
FT replace(105494,C)
FT /*tag= bc
FT /standard name= "Single nucleotide polymorphism"
FT replace(105911,T)
FT /*tag= bd
FT /standard name= "Single nucleotide polymorphism"
FT 108868..108959
FT /*tag= be
FT 108960..110489
FT /*tag= bf
FT replace(110376,C)
FT /*tag= bg
FT /standard name= "Single nucleotide polymorphism"
FT 110490..110579
FT /*tag= bh
FT 110580..115705
FT /*tag= bi
FT replace(111250,A)
FT /*tag= bj
FT /standard name= "Single nucleotide polymorphism"
FT replace(111990,C)
FT /*tag= bk
FT /standard name= "Single nucleotide polymorphism"
FT replace(112748,A)
FT /*tag= bl
FT /standard name= "Single nucleotide polymorphism"
FT replace(112783,G)
FT /*tag= bm
FT /standard name= "Single nucleotide polymorphism"
FT replace(114256,A)
FT /*tag= bn
FT /standard name= "Single nucleotide polymorphism"
FT 115706..115863
FT /*tag= bo
FT 115864..118931
FT /*tag= bp
```

Query Match 76.8%; Score 19.2; DB 10; Length 202001;
Best Local Similarity 87.5%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 TGAAGAAATTCAGTTTCATAGCTTG 24
DB 174078 TGAAGAAATTCATTTCAAAGCTTG 174055
```

RESULT 10

```
ADP92027
ID ADP92027 standard; cDNA; 235 BP.
XX ADP92027;
XX 09-SEP-2004 (first entry)
XX Cotton expressed sequence tag, EST, #1038.
XX Cotton; ss; EST; expressed sequence tag; plant; plant protection;
XX plant improvement; marker-assisted breeding.
XX Gossypium hirsutum; variety Nucleotid33B.
XX US2004123338-A1.
XX 24-JUN-2004.
XX 08-DEC-2000; 2000US-00732627.
XX 10-DEC-1999; 99US-0170255P.
XX (FINC/) FINCHER K L.
XX Fincher XL;
XX WPI; 2004-479807/45.
XX New substantially purified nucleic acid molecule that encodes a cotton
XX protein or its fragment, useful as molecular tool for the targeting and
XX isolation of novel genes for plant protection and improvement.
XX Claim 1; SEQ ID NO 1038; 30pp; English.
XX The invention relates to a substantially purified nucleic acid molecule
XX that encodes a cotton protein or its fragment comprising an EST
XX (expressed sequence tag) appearing as ADP9090-ADP95919. Also included
XX are a substantially purified cotton protein or its fragment encoded by a
XX nucleic acid molecule above and a transformed plant (having a nucleic
XX acid molecule which comprises: an exogenous promoter region which
XX functions in a plant cell to cause the production of a mRNA molecule; a
XX structural nucleic acid molecule comprising one of the ESTs or their
XX complements; a 3' non-translated sequence that functions in the plant
XX cell to cause termination of transcription and addition of polyadenylated
XX ribonucleotides to a 3' end of the mRNA molecule). The ESTs are useful as
XX a molecular tool for the targeting and isolation of novel genes for plant
XX protection and improvement. The ESTs are useful for developing new
XX strategies for understanding critical plant developmental and metabolic
XX pathways, for isolating genes and promoters, for identifying and mapping
XX the genes involved in developmental and metabolic pathways, and for
XX determining gene function. The cotton nucleic acid molecules are useful
XX as molecular tags to isolate genetic regions, isolate genes, map genes,
XX and determine gene function. The nucleic acid molecules are useful for
XX determining if genes are members of a particular gene family and for use
XX in marker-assisted breeding programs. The present sequence is one of the
XX 4930 cotton ESTs of the invention. NOTE: The sequences are not displayed
XX in the specification but are available in electronic format from the
XX USPTO at seqdata.uspto.gov/sequence.html?DocID=20040123338.
```

SQ Sequence 235 BP; 66 A; 40 C; 60 G; 69 T; 0 U; 0 Other;

Query Match 76.0%; Score 19; DB 12; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TGAAGAAATTCAGTTTCATA 19
```

```

Db      179 TGAAGAAATTCAGTTCATA 197
|||||
RESULT 11
ACH97547/c
ID ACH97547 standard; DNA; 1227 BP.
XX AC ACH97547;
XX DT 29-JUL-2004 (first entry)
XX DE Klebsiella pneumoniae polynucleotide seqid 3342.
XX KW Recombinant expression vector; transcription regulatory element;
XX KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX OS Klebsiella pneumoniae.
XX PN US6610836-B1.
XX PD 26-AUG-2003.
XX PF 27-JAN-2000; 2000US-00489039.
XX PR 29-JAN-1999; 99US-0117747P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton GL, Osborne M;
XX DR WPI; 2003-895346/82.
XX DR P-PSDB; ABO63996.
XX PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
XX PT preparing a vaccine composition against Klebsiella pneumoniae.
XX PS Disclosure; SEQ ID NO 3342; 932pp; English.
XX CC The invention describes a new isolated nucleic acid encoding a Klebsiella
XX CC pneumoniae polypeptide. Also described are: a recombinant expression
XX CC vector comprising the nucleic acid, operably linked to a transcription
XX CC regulatory element; and a cell comprising the recombinant expression
XX CC vector. The nucleic acid is useful for preparing a vaccine composition
XX CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
XX CC pneumoniae polypeptide of the invention
XX SQ Sequence 1227 BP; 222 A; 351 C; 360 G; 294 T; 0 U; 0 Other;

Query Match 75.2%; Score 18.8; DB 11; Length 1227;
Best Local Similarity 90.9%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCT 22
|||||
Db 937 TGATGAATCCAGTTCATAGCT 916

RESULT 12
ADQ97167/c
ID ADQ97167 standard; DNA; 176771 BP.
XX AC ADQ97167;
XX DT 07-OCT-2004 (first entry)
XX DE Human cancer associated sequence HD2-08-009, SEQ ID 143.
XX KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX OS Homo sapiens.
XX PS WO2004060304-A2.

```

```

XX 22-JUL-2004.
XX 22-DEC-2003; 2003WO-US041389.
XX 27-DEC-2002; 2002US-00330773.
XX (SAGR-) SAGRES DISCOVERY INC.
XX Morris DW, Malandro MS;
XX WPI; 2004-543781/52.
XX New isolated cancer associated nucleic acids comprising at least 10
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX PT cancers such as leukemia and lymphoma.
XX PS Claim 1; SEQ ID NO 143; 199pp; English.
XX CC The present invention relates to cancer associated sequences (ADQ97025-
XX CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 176771 BP; 47665 A; 34089 C; 38342 G; 56655 T; 0 U; 20 Other;

Query Match 75.2%; Score 18.8; DB 12; Length 176771;
Best Local Similarity 90.9%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCT 22
|||||
Db 100466 TGAAGAAATTCATTATAGCT 100445

RESULT 13
ABZ14865
ID ABZ14865 standard; DNA; 4212 BP.
XX AC ABZ14865;
XX DT 21-JAN-2003 (first entry)
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2670.
XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX OS Arabidopsis thaliana.
XX PN WO200216655-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US026685.
XX PR 24-AUG-2000; 2000US-0227866P.
XX PR 26-JAN-2001; 2001US-0264647P.
XX PR 22-JUN-2001; 2001US-0300111P.
XX PA (SCRI ) SCRIPPS RES INST.
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PI Harper JF, Kreps J, Wang X, Zhu T;
XX WPI; 2002-304127/34.
XX PT Identifying a stress condition to which a plant cell has been exposed and
XX PT producing plants with increased tolerance to these abiotic stresses.
XX PS Claim 144; SEQ ID NO 2670; 577pp + Sequence Listing; English.

```

CC The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising: (a) contacting nucleic acid with an
 CC representative of expressed polynucleotides in the plant cell with an
 CC array or probes representative of the plant cell genome; and (b)
 CC detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
 CC in methods of the invention. Note: The sequence data for this patent is
 CC not represented in the printed specification but is based on sequence
 CC information supplied to Derwent by the European Patent Office
 XX
 SQ Sequence 4212 BP; 1235 A; 892 C; 963 G; 1122 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 4212;
 Best Local Similarity 84.0%; Pred. No. 2.6e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCTTGT 25
 Db 2453 TGAAGAAAGTTAGCTCATACCTTGT 2477

RESULT 14
 ACN44212
 ID ACN44212 standard; DNA; 116858 BP.

XX ACN44212;

DT 18-NOV-2004 (first entry)

DE Mouse genomic sequence MCG15964.

XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.

XX Mus musculus.

XX WO2003073826-A2.

PD 12-SEP-2003.

PF 28-FEB-2003; 2003WO-US006235.

PR 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

PA Morris DW;

DR WPI; 2003-328604/31.

XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 PT comprises a nucleotide sequence.

PS Claim 1; SEQ ID NO 547; Opp; English.

XX The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CA coding
 CC sequence. Note: This patent is an equivalent to basic patent
 CC US2002182586A1, for which no sequence data was published
 XX

SQ Sequence 116858 BP; 32402 A; 22234 C; 23982 G; 34772 T; 0 U; 3468 Other;

Query Match 74.4%; Score 18.6; DB 11; Length 116858;
 Best Local Similarity 84.0%; Pred. No. 3.5e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCTTGT 25
 Db 17178 TGAAGAAATTCATATATTTTGT 17202

RESULT 15
 ADK54409/c

ID ADK54409 standard; DNA; 978 BP.

XX ADK54409;

DT 06-MAY-2004 (first entry)

XX Plant DNA sequence which confers altered metabolic characteristic #1792.

XX altered metabolic characteristic; plant; acid metabolism;

KW alcohol metabolism; fatty acid metabolism;

KW branched fatty acid metabolism; alkaloid metabolism;

KW amino acid metabolism; ester metabolism; glyceride metabolism;

KW phenolic metabolism; carbohydrate metabolism; sterol metabolism;

KW terpene metabolism; isoprenoid metabolism; alkene metabolism;

KW alkyne metabolism; hydrocarbon metabolism; ketone metabolism;

KW quinone metabolism; disease resistance; gene shuffling; sexual PCR; ds.

XX Unidentified.

XX WO2003020936-A1.

XX 13-MAR-2003.

XX 30-AUG-2002; 2002WO-US027884.

XX 31-AUG-2001; 2001US-0316471P.

XX (DOWC) DOW CHEM CO.

XX (DOWC) DOW AGROSCIENCES LLC.

XX Weglarz T, Gachotte D, Blakeslee B, McCreary DA, Pell RJ;

PI Oriedo JVB, Crosley R, Reddy AS, Shukla V, Larrinua I, Miller BA;

XX WPI; 2003-313091/30.

XX Novel genes that confer altered metabolic characteristics in Nicotiana
 PT benthamiana plants, useful for altering the levels of metabolites e.g.
 PT acids, fatty acids, amino acids, carbohydrates, hydrocarbons and sterols.

PS Claim 1; SEQ ID NO 1792; 2576pp; English.

XX The invention comprises DNA sequences which confer an altered metabolic
 CC characteristic when they are expressed in a plant. The DNA sequences of
 CC the invention are useful for producing plants with an altered metabolic
 CC characteristic, such as: altered acid metabolism, alcohol metabolism,
 CC fatty acid metabolism, branched fatty acid metabolism, alkaloid or other
 CC base metabolism, altered amino acid metabolism, altered ester metabolism,
 CC altered glyceride metabolism, altered phenolic metabolism, altered
 CC carbohydrate metabolism, altered sterol, oxygenated terpene, or
 CC isoprenoid metabolism, alkene or alkyne metabolism, hydrocarbon
 CC metabolism, ketone or quinone metabolism. The DNA sequences of the
 CC invention may be used to provide disease resistance in a plant and gene
 CC shuffling or sexual PCR procedures. The present nucleic acid represents a
 CC DNA sequence of the invention.

SQ Sequence 978 BP; 358 A; 146 C; 164 G; 310 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 10; Length 978;
 Best Local Similarity 87.0%; Pred. No. 3.3e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AAGAAATTCAGTTCATAGCTTGT 25
|||
Db 801 AATAAATTCAGTTCAGTTCATAGCTTGT 779

Search completed: August 13, 2005, 04:14:41
Job time : 206.266 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 03:33:32 ; Search time 61.4362 Seconds
(without alignments)
665.844 Million cell updates/sec

Title: US-10-673-854-3

Perfect score: 25

Sequence: 1 tgaagaattcagttcagttcgtgt 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfileseq.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|-------------------|
| C 1 | 19.8 | 79.2 | 44848 | US-09-435-739-42 | Sequence 42, Appl |
| C 2 | 19.8 | 79.2 | 44848 | US-09-988-113-42 | Sequence 42, Appl |
| C 3 | 19.2 | 76.8 | 202001 | US-09-734-674-3 | Sequence 3, Appli |
| C 4 | 18.8 | 75.2 | 1227 | US-09-489-039A-3342 | Sequence 3342, Ap |
| C 5 | 18.2 | 72.8 | 29326 | US-09-949-016-15356 | Sequence 15356, A |
| C 6 | 18.2 | 72.8 | 29720 | US-09-949-016-16321 | Sequence 16321, A |
| 7 | 18.2 | 72.8 | 126176 | US-09-949-016-16137 | Sequence 16137, A |
| 8 | 18.2 | 72.8 | 126176 | US-09-949-016-16138 | Sequence 16138, A |
| C 9 | 18.2 | 72.8 | 223471 | US-09-949-016-12387 | Sequence 12387, A |
| C 10 | 18.2 | 72.8 | 223471 | US-09-949-016-12724 | Sequence 12724, A |
| C 11 | 18.2 | 72.8 | 223471 | US-09-949-016-12725 | Sequence 12725, A |
| 12 | 18 | 72.0 | 447 | US-09-621-976-17257 | Sequence 17257, A |
| 13 | 17.8 | 71.2 | 601 | US-09-949-016-157433 | Sequence 157433, |
| 14 | 17.8 | 71.2 | 601 | US-09-949-016-157540 | Sequence 157540, |
| C 15 | 17.8 | 71.2 | 1325 | US-09-710-279-999 | Sequence 999, App |
| C 16 | 17.8 | 71.2 | 1485 | US-09-134-001C-1339 | Sequence 1339, Ap |
| 17 | 17.8 | 71.2 | 2978 | US-09-710-279-4099 | Sequence 4099, Ap |
| 18 | 17.8 | 71.2 | 3745 | US-09-710-279-4045 | Sequence 4045, Ap |
| 19 | 17.6 | 70.4 | 41988 | US-09-949-016-14501 | Sequence 14501, A |
| 20 | 17.6 | 70.4 | 77618 | US-09-949-016-11768 | Sequence 11768, A |
| 21 | 17.6 | 70.4 | 119153 | US-09-949-016-12378 | Sequence 12378, A |
| 22 | 17.4 | 69.6 | 601 | US-09-949-016-179792 | Sequence 179792, |
| 23 | 17.4 | 69.6 | 601 | US-09-949-016-179793 | Sequence 179793, |
| 24 | 17.4 | 69.6 | 390416 | US-09-348-352-1292 | Sequence 1292, Ap |
| 25 | 17.2 | 68.8 | 192 | US-09-328-352-1292 | Sequence 1292, Ap |
| 26 | 17.2 | 68.8 | 601 | US-09-949-016-93041 | Sequence 93041, A |
| 27 | 17.2 | 68.8 | 601 | US-09-949-016-93042 | Sequence 93042, A |

| | | | | | |
|------|------|------|--------|----------------------|-------------------|
| C 28 | 17.2 | 68.8 | 601 | US-09-949-016-153121 | Sequence 153121, |
| C 29 | 17.2 | 68.8 | 601 | US-09-949-016-153122 | Sequence 153122, |
| C 30 | 17.2 | 68.8 | 601 | US-09-949-016-153123 | Sequence 153123, |
| C 31 | 17.2 | 68.8 | 601 | US-09-949-016-177341 | Sequence 177341, |
| 32 | 17.2 | 68.8 | 601 | US-09-949-016-177342 | Sequence 177342, |
| 33 | 17.2 | 68.8 | 601 | US-09-949-016-177343 | Sequence 177343, |
| 34 | 17.2 | 68.8 | 601 | US-09-949-016-177344 | Sequence 177344, |
| 35 | 17.2 | 68.8 | 2521 | US-09-620-312D-1015 | Sequence 1015, Ap |
| 36 | 17.2 | 68.8 | 2520 | US-09-620-312D-1014 | Sequence 1014, Ap |
| 37 | 17.2 | 68.8 | 13438 | US-09-949-016-16808 | Sequence 16808, A |
| 38 | 17.2 | 68.8 | 35058 | US-09-949-016-12607 | Sequence 12607, A |
| 39 | 17.2 | 68.8 | 35059 | US-09-949-016-13831 | Sequence 13831, A |
| C 40 | 17.2 | 68.8 | 74177 | US-09-949-016-11988 | Sequence 11988, A |
| C 41 | 17.2 | 68.8 | 74177 | US-09-949-016-17388 | Sequence 17388, A |
| C 42 | 17.2 | 68.8 | 83697 | US-09-949-016-16040 | Sequence 16040, A |
| C 43 | 17.2 | 68.8 | 126468 | US-09-949-016-14418 | Sequence 14418, A |
| 44 | 17.2 | 68.8 | 131724 | US-09-949-016-12893 | Sequence 12893, A |
| C 45 | 17.2 | 68.8 | 176006 | US-09-949-016-16804 | Sequence 16804, A |

ALIGNMENTS

RESULT 1

US-09-435-739-42/c

; Sequence 42, Application US/09435739

; Patent No. 6664105

; GENERAL INFORMATION:

; APPLICANT: Pecker, Iris

; APPLICANT: Vlodavsky, Israel

; APPLICANT: Feinstein, Elena

; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE HAVING HEPARANASE ACTIVITY

; TITLE OF INVENTION: EXPRESSION OF SAME IN GENETICALLY MODIFIED CELLS

; FILE REFERENCE: 00/20454

; CURRENT APPLICATION NUMBER: US/09/435,739

; CURRENT FILING DATE: 2001-06-05

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 42

; LENGTH: 44848

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-435-739-42

Query Match 79.2%; Score 19.8; DB 4; Length 44848;
Best Local Similarity 91.3%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAAGAAATTCAGTTTCATAGCTTG 24

Db 30803 GAAGAAATTCAGTTTCATAGCTTG 30781

RESULT 2

US-09-988-113-42/c

; Sequence 42, Application US/09988113

; Patent No. 6790658

; GENERAL INFORMATION:

; APPLICANT: Pecker, Iris

; APPLICANT: Vlodavsky, Israel

; APPLICANT: Feinstein, Elena

; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE HAVING HEPARANASE ACTIVITY

; TITLE OF INVENTION: EXPRESSION OF SAME IN GENETICALLY MODIFIED CELLS

; FILE REFERENCE: 01/22781

; CURRENT APPLICATION NUMBER: US/09/988,113

; CURRENT FILING DATE: 2001-11-19

; PRIOR APPLICATION NUMBER: US 09/776,874

; PRIOR FILING DATE: 2001-02-06

; PRIOR APPLICATION NUMBER: US09/258,892

; PRIOR FILING DATE: 1999-03-01

; PRIOR APPLICATION NUMBER: PCT/US98/17954

; PRIOR FILING DATE: 1998-08-31

; PRIOR APPLICATION NUMBER: US 09/109,386

```
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 08/922,170
; PRIOR FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 44848
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-988-113-42

Query Match      79.2%; Score 19.8; DB 4; Length 44848;
Best Local Similarity 91.3%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  2 GAAGAAATTCAGTTCATAGCTTG 24
    |||||
Db  30803 GAAGAAATTCAGTTCATAGCTTG 30781

RESULT 3
US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. 6498022
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

Query Match      76.8%; Score 19.2; DB 4; Length 202001;
Best Local Similarity 87.5%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  1 TGAAGAAATTCAGTTCATAGCTTG 24
    |||||
Db  174078 TGAAGAAATTCATTCAGTTCATAGCTTG 174055

RESULT 4
US-09-489-039A-3342/c
; Sequence 3342, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3342
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3342

; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 08/922,170
; PRIOR FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 44848
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-988-113-42

Query Match      75.2%; Score 18.8; DB 4; Length 1227;
Best Local Similarity 90.9%; Pred. No. 58;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1 TGAAGAAATTCAGTTCATAGCT 22
    |||||
Db  937 TGATGAATCCAGTTCATAGCT 916

RESULT 5
US-09-949-016-15356/c
; Sequence 15356, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15356
; LENGTH: 29326
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15356

Query Match      72.8%; Score 18.2; DB 4; Length 29326;
Best Local Similarity 87.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  2 GAAGAAATTCAGTTCATAGCTTG 24
    |||||
Db  11592 GAAGAAATTCAGTTCGTGCTTG 11570

RESULT 6
US-09-949-016-16521/c
; Sequence 16521, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16521
; LENGTH: 29720
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16521

Query Match      72.8%; Score 18.2; DB 4; Length 29720;
Best Local Similarity 87.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  2 GAAGAAATTCAGTTCATAGCTTG 24
```

Db 4957 GAAGGAATTCAGTTCATGCTTG 4935
|||||

RESULT 7
US-09-949-016-16137
; Sequence 16137, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16137
; LENGTH: 126176
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16137

Query Match 72.8%; Score 18.2; DB 4; Length 126176;
Best Local Similarity 87.0%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATGCTTG 24
|||||

Db 44879 GAAGAAATTCAGTTCCTAGCTGG 44901
|||||

RESULT 8
US-09-949-016-16138
; Sequence 16138, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16138
; LENGTH: 126176
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16138

Query Match 72.8%; Score 18.2; DB 4; Length 126176;
Best Local Similarity 87.0%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATGCTTG 24
|||||

Db 44879 GAAGAAATTCAGTTCCTAGCTGG 44901
|||||

RESULT 9

US-09-949-016-12387/c
; Sequence 12387, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12387
; LENGTH: 223471
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(223471)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12387

Query Match 72.8%; Score 18.2; DB 4; Length 223471;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATGCTTG 24
|||||

Db 222918 GAAGAAATTCAGTTCATGCTTG 222896
|||||

RESULT 10
US-09-949-016-12724/c
; Sequence 12724, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12724
; LENGTH: 223471
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(223471)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12724

Query Match 72.8%; Score 18.2; DB 4; Length 223471;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATGCTTG 24
|||||

Db 222918 GAAGAAATTCAGTTCATGCTTG 222896
|||||

RESULT 11
US-09-949-016-12725/c
; Sequence 12725, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12725
; LENGTH: 223471
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(223471)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12725

Query Match 72.8%; Score 18.2; DB 4; Length 223471;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24
|||||
Db 222918 GAAGAAATTCAGTTCATAGCTTG 222896

RESULT 12
US-09-621-976-17257
; Sequence 17257, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 17257
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 371,395,397,407,410
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-17257

Query Match 72.0%; Score 18; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCAT 18
|||||
Db 287 TGAAGAAATTCAGTTCAT 304

RESULT 13
US-09-949-016-157433
; Sequence 157433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 157433
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-157433

Query Match 71.2%; Score 17.8; DB 4; Length 601;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24
|||||
Db 290 GAAGAAATTCATCTCCTAGCTGG 312

RESULT 14
US-09-949-016-157540
; Sequence 157540, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 157540
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-157540

Query Match 71.2%; Score 17.8; DB 4; Length 601;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24
|||||
Db 290 GAAGAAATTCATCTCCTAGCTGG 312

RESULT 15
US-09-710-279-999/c
; Sequence 999, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 999
; LENGTH: 1325
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-999

Query Match 71.2%; Score 17.8; DB 4; Length 1325;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 GAAATTCAGTTCATAGCTTCT 25
| | | | | | | | | | | | | | | | | | | | | |
Db 1305 GAAATTCAGTTCATAGCTTCT 1285

Search completed: August 13, 2005, 06:48:53
Job time : 65.4362 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 03:26:33 ; Search time 1606.65 Seconds
(without alignments)
592.293 Million cell updates/sec

Title: US-10-673-854-3

Perfect score: 25
Sequence: 1 tgaagaaattcagttcatagctgt 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 20.2 | 80.8 | 187 | 6 | CD025522 NXSI_060 |
| 2 | 20.2 | 80.8 | 623 | 7 | CO198363 GRC1_13.D |
| 3 | 20.2 | 80.8 | 780 | 7 | CF689706 RNCNT1_45 |
| 4 | 20.2 | 80.8 | 902 | 7 | CO226474 WS01029.B |
| 5 | 19.8 | 79.2 | 280 | 5 | BQ236799 TAE05026F |
| 6 | 19.8 | 79.2 | 395 | 6 | CA707706 wdk2c.pk0 |
| 7 | 19.8 | 79.2 | 546 | 6 | CD895321 G174_001I |
| 8 | 19.8 | 79.2 | 581 | 2 | BE402372 CSB007C06 |
| 9 | 19.8 | 79.2 | 581 | 5 | BQ607758 BRY_3653 |
| 10 | 19.8 | 79.2 | 634 | 4 | BQ250598 BQ250598 |
| 11 | 19.8 | 79.2 | 636 | 6 | CD896013 G174_101J |
| 12 | 19.8 | 79.2 | 643 | 6 | CD453897 WHE0902.E |
| 13 | 19.8 | 79.2 | 773 | 5 | BQ804516 WHE3555.F |
| 14 | 19.8 | 79.2 | 783 | 8 | BZ186617 CH230-340 |
| 15 | 19.8 | 79.2 | 863 | 9 | CC495993 CH240_331 |
| 16 | 19.8 | 79.2 | 1002 | 9 | CNS05L98 |
| 17 | 19.4 | 77.6 | 710 | 7 | CK367386 AGENCOURT |
| 18 | 19.2 | 76.8 | 506 | 7 | CG639240 OST369798 |
| 19 | 19.2 | 76.8 | 536 | 7 | CN518432 G00094.B3 |
| 20 | 19.2 | 76.8 | 588 | 8 | AZ229308 RRC1-23-5 |
| 21 | 19.2 | 76.8 | 599 | 8 | AQ987207 RRC1-23-3 |
| 22 | 19.2 | 76.8 | 662 | 5 | BQ139159 NF011D06P |
| 23 | 19.2 | 76.8 | 689 | 8 | B92536 CIT-HSP-217 |
| 24 | 19.2 | 76.8 | 725 | 9 | CE361033 tigr-gss- |

| | | | | | | | | |
|------|------|------|------|---|----------|-----------|-----------|-----|
| C 25 | 19.2 | 76.8 | 775 | 7 | CN986329 | CN986329 | 61938 | 125 |
| C 26 | 19.2 | 76.8 | 810 | 5 | BG760956 | BG760956 | 60217652 | |
| C 27 | 19.2 | 76.8 | 961 | 4 | BU333932 | BU333932 | 603499379 | |
| C 28 | 19.2 | 76.8 | 963 | 9 | CL053389 | CL053389 | CH216-77J | |
| C 29 | 19.2 | 76.8 | 1291 | 9 | AG435478 | Mus muscu | | |
| C 30 | 18.8 | 75.2 | 231 | 8 | BZ416853 | if71009.g | | |
| C 31 | 18.8 | 75.2 | 529 | 6 | CD312483 | StrPu691. | | |
| C 32 | 18.8 | 75.2 | 544 | 8 | AZ007802 | RFC1-23-3 | | |
| C 33 | 18.8 | 75.2 | 619 | 4 | BG085680 | H3116C06- | | |
| C 34 | 18.8 | 75.2 | 699 | 8 | AZ106414 | RPC1-23-3 | | |
| C 35 | 18.8 | 75.2 | 704 | 1 | AL698618 | DKFZ0686K | | |
| C 36 | 18.8 | 75.2 | 726 | 9 | CE132344 | tigr-gss- | | |
| C 37 | 18.8 | 75.2 | 1058 | 9 | CN905MCC | | | |
| C 38 | 18.8 | 75.2 | 2625 | 9 | CL960736 | OsfCC037 | | |
| C 39 | 18.6 | 74.4 | 143 | 6 | CD313427 | StrPu621. | | |
| C 40 | 18.6 | 74.4 | 241 | 9 | CL325373 | RPC144_24 | | |
| C 41 | 18.6 | 74.4 | 275 | 8 | AQ103010 | HS_3059_A | | |
| C 42 | 18.6 | 74.4 | 349 | 5 | BQ32895 | QV0-ET014 | | |
| C 43 | 18.6 | 74.4 | 354 | 2 | AW756746 | B126901.Y | | |
| C 44 | 18.6 | 74.4 | 383 | 8 | AQ238508 | RPC111-63 | | |
| C 45 | 18.6 | 74.4 | 425 | 4 | BM176564 | TJESTzya9 | | |

ALIGNMENTS

RESULT 1
LOCUS CD025522 187 bp mRNA linear EST 07-MAY-2003
DEFINITION NXSI_060_F12_F NXSI (Nsf Xylem Side wood Inclined) Pinus taeda cDNA
clone NXSI_060_F12_5', mRNA sequence.
ACCESSION CD025522
VERSION CD025522.1 GI:30364172
KEYWORDS EST.
SOURCE Pinus taeda (loblolly pine)
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 187)
Sederoff, R.
Molecular Basis of Wood Formation in the Pine Megagenome
Unpublished (2000)
Contact: Sederoff, Ron
Forest Biotechnology
North Carolina State University
840 Main Campus Drive, Centennial Campus, Campus Box 7247, Raleigh,
NC 27695, USA
Tel: 919 515 7800
Fax: 919 515 7801
Email: ron.sederoff@ncsu.edu, jerril.johnson@ncsu.edu
Please see <http://web.abu.umn.edu/biodata/nsfpine/> for further
information.
Seq primer: T3.

FEATURES
Location/Qualifiers
1..187
/organism="Pinus taeda"
/mol_type="mRNA"
/strain="Coastal plain loblolly pine from North Carolina"
/db_xref="taxon:3352"
/clone="NXSI_060_F12"
/tissue_type="Xylem"
/cell_type="Side"
/dev_stage="Juvenile"
/lab_host="XL1-Blue"
/note="Vector: BlueScript SK; Site 1: Eco RI; Site 2:
XhoI; The library is from early (spring) wood, taken from
three six-year old trees (three different genotypes), in
the juvenile phase. These trees were induced to form side
wood by bending to a 45 degree angle and tying them to the
ground. Differentiating xylem was harvested from the sides
of the inclined stems, and a mixture of all three
genotypes was used for the library. oligo-dT primed cDNA

was directionally cloned into the EcoRI-XhoI Bluescript SK vector arms. NOTE: The sequences contain a 'cDNA adapter' between the EcoRI site and the start of the EST. The adapter sequence is 'AATTCGCCAG'."

ORIGIN
Query Match 80.8%; Score 20.2; DB 6; Length 187;
Best Local Similarity 88.0%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTTCATAGCTTGT 25
|||||
Db 125 TGATGAATTCAGTTTCATACCATGT 149

RESULT 2
CO198363
LOCUS
DEFINITION
cDNA clone GEO1_13_D06_g1_A029 Root gravitropism April 2003 test Pinus taeda
CO198363
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pinus taeda (loblolly pine)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE
AUTHORS
Pratt,L., Cordonnier-Pratt,M.-M., Lorenz,W.W., Zimmermann,C. and Dean,J.F.D.
TITLE
An EST database from gravitropically stimulated loblolly pine (Pinus taeda) roots
JOURNAL
COMMENT
Other_ESTs: GEO1_13_D06_g1_A029
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

RNA prepared and library constructed by W. Walter Lorenz (School of Forest Resources, University of Georgia); plant material prepared by Craig Zimmermann (School of Forest Resources, University of Georgia) using rooted cuttings provided by the Forest Biology Research Cooperative (FBRC) and the CCLONES project a the University of Florida; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.
Seq primer: M13-21 (TGTAACGACGCCAGT)
POLYA=Yes.

FEATURES
source
1..623
/organism="Pinus taeda"
/mol_type="mRNA"
/strain="3 CCLONES"
/db_xref="taxon:3352"
/clone="GEO1_13_D06_A029"
/lab_host="DH10B-T1 phage-resistant E. coli"
/note="Vector: pSL180; Site 1: EcoRI; Site 2: XhoI; The library was prepared from polyA+ RNA from the roots of 1-year-old loblolly pine (Pinus taeda) cuttings that were rooted and then planted in washed sand. The rooted cuttings were maintained for 27 days (April 2003 harvest) under ambient conditions in a local greenhouse. They were kept on a weekly regimen of 0.5x nutrient-complete Hoagland's solution and supplemented with additional water sufficient to maintain a 15% soil moisture content. Twenty-four hours (24 h) prior to harvesting roots for mRNA preparation, the potted trees were tipped 90 degree

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 623;
Best Local Similarity 88.0%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTTCATAGCTTGT 25
|||||
Db 180 TGATGAATTCAGTTTCATACCATGT 204

RESULT 3
CF669706
LOCUS
DEFINITION
RTCN1_45_C06_g1_A029 Root control Pinus taeda cDNA clone
CF669706
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pinus taeda (loblolly pine)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE
AUTHORS
Pratt,L., Cordonnier-Pratt,M.-M., Lorenz,W.W., Zimmermann,C. and Dean,J.F.D.
TITLE
An EST database from untreated loblolly pine (Pinus taeda) roots
JOURNAL
COMMENT
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

RNA prepared and library constructed by W. Walter Lorenz (School of Forest Resources, University of Georgia); plant material prepared by Craig Zimmermann (School of Forest Resources, University of Georgia) using rooted cuttings provided by the Forest Biology Research Cooperative (FBRC) and the CCLONES project a the University of Florida; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.
Seq primer: JENREV (CAGGAACAGCTATGACC).

FEATURES
source
1..780
/organism="Pinus taeda"
/mol_type="mRNA"
/strain="3 CCLONES"
/db_xref="taxon:3352"
/clone="RTCN1_45_C06_A029"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Root control"
/note="Organ: root; Vector: pSL1180; Site 1: EcoRI; Site 2: XhoI; The library was prepared from polyA+ RNA from the roots of 1-year-old loblolly pine (Pinus taeda) cuttings that were rooted and then planted in washed sand. Just before harvesting roots for RNA isolation, the rooted cuttings were maintained for 27 days (April 2003) under ambient conditions in a local greenhouse. They were kept on a weekly regimen of 0.5x nutrient-complete Hoagland's solution and supplemented with additional water sufficient to maintain a 15% soil moisture content. Double-stranded cDNA was cloned unidirectionally into pSL180. Inserts can be excised with EcoRI (5' end) and XhoI (3' end)."

ORIGIN

to lay the pots on their sides to induce a gravitropic response. Double-stranded cDNA was cloned unidirectionally into pSL1180. Inserts can be excised with EcoRI (5' end) and XhoI (3' end)."

[illegible]

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.

1 (bases 1 to 395)
Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,
Miao, G., Caraher, N. and Hanafey, M.K.

Unpublished (2002)
DuPont Wheat cDNA Sequence

Contact: Scott V. Tingey

Crop Genetics

E. I. Dupont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

Tel: 302-631-2602

Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com

Seq primer: M13.

Location/Qualifiers

1. .395

/organism="Triticum aestivum"

/mol_type="mRNA"

/db_xref="taxon:4565"

/clone="wdk2c.pk006.b16"

/tissue_type="kernel"

/clone_lib="wdk2c"

/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
days after anthesis."

ORIGIN

Query Match 79.2%; Score 19.8; DB 6; Length 395;
Best Local Similarity 91.3%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24

|||||

Db 327 GAAGAAATTCAGTTCCTAGCATG 305

RESULT 7

CD895321/c
LOCUS 546 bp mRNA linear EST 14-JUL-2003
DEFINITION G174.00115F010514 G174 Triticum aestivum cDNA clone G17400115,
mRNA sequence.

ACCESSION CD895321.1 GI:32667779

VERSION EST.

KEYWORDS Triticum aestivum (bread wheat)

SOURCE Triticum aestivum

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.

1 (bases 1 to 546)

/dev_stage="8-12 days post anthesis"

/lab_host="Escherichia coli SOLR"

/clone_lib="ITEC CSB Wheat Endosperm Library"

/note="Vector: Lambda Zap/Bluescript; Site 1: XhoI;
Site 2: EcoRI; Plants grown in Phytotron with 18C/13C
(day/night) 16 hour light. M13 Reverse sequencing primer
used. 1.0 Kbp average insert size."

ORIGIN

Query Match 79.2%; Score 19.8; DB 2; Length 581;
Best Local Similarity 91.3%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24

|||||

Db 575 GAAGAAATTCAGTTCCTAGCATG 553

FEATURES

source

1. .546

/organism="Triticum aestivum"

/mol_type="mRNA"

/cultivar="recital"

/db_xref="taxon:4565"

/clone="G17400115"

/tissue_type="grain (174 degrees per day after
pollination)"

/clone_lib="G174"

ORIGIN

Query Match 79.2%; Score 19.8; DB 6; Length 546;
Best Local Similarity 91.3%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24

|||||

Db 486 GAAGAAATTCAGTTCCTAGCATG 464

RESULT 8

BE402372/c

LOCUS 581 bp mRNA linear EST 21-JUL-2000

DEFINITION CSB007C06F90908 ITEC CSB Wheat Endosperm Library Triticum aestivum

CDNA clone CSB007C06, mRNA sequence.

ACCESSION BE402372.1 GI:9361840

VERSION EST.

KEYWORDS Triticum aestivum (bread wheat)

SOURCE Triticum aestivum

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.

1 (bases 1 to 581)

/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
days after anthesis."

ORIGIN

Query Match 79.2%; Score 19.8; DB 6; Length 546;
Best Local Similarity 91.3%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24

|||||

Db 327 GAAGAAATTCAGTTCCTAGCATG 305

RESULT 9

BQ607758/c

LOCUS 581 bp mRNA linear EST 25-JUN-2002

DEFINITION BRY 3653 wheat EST endosperm library Triticum aestivum cDNA 5',
mRNA sequence.

ACCESSION BQ607758

```

VERSION BQ607758.1 GI:21557087
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 581)
AUTHORS Clarke, B., Lambrecht, M. and Rhee, S.Y.
TITLE Arabidopsis genomic information for interpreting wheat EST
sequences
JOURNAL Funct. Integr. Genomics 3 (1-2), 33-38 (2003)
MEDLINE 22478026
PUBMED 12590341
COMMENT Contact: Lambrecht M
The Arabidopsis Information Resource
Carnegie Institution of Washington, Dept. of Plant Biology
260 Panama Street, Stanford, CA 94305, USA
Tel: 1 650 325 1521 x 251
Fax: 1 650 325 3748
Email: rhee@acoma.stanford.edu.
FEATURES
source
1..581
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wyuana"
/db_xref="taxon:4565"
/tissue_type="endosperm"
/dev_stage="developing endosperm tissue 8, 10 and 12 DPA
(days post anthesis)"
/clone_lib="wheat EST endosperm library"
ORIGIN
Query Match 79.2%; Score 19.8; DB 5; Length 581;
Best Local Similarity 91.3%; Pred. NO. 3.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAAGAAATTCAGTTCATAGCTTG 24
|||||
Db 575 GAAGAAATTCAGTTCCTAGCATG 553

RESULT 10
BJ250598 634 bp mRNA linear EST 05-APR-2002
LOCUS BJ250598 Y. Ogiwara unpublished cDNA library, wh_f Triticum
DEFINITION aestivum cDNA clone wh16f11 3', mRNA sequence.
ACCESSION BJ250598
VERSION BJ250598.1 GI:20060583
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 634)
AUTHORS Ogiwara, Y. and Murai, K.
TITLE Expressed genes in Triticum aestivum
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..634
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="wh16f11"

```

```

/tissue_type="spike at flowering date"
/dev_stage="Peekes" scale 10.5.1"
/clone_lib="Y. Ogiwara unpublished cDNA library, wh_f"
ORIGIN
Query Match 79.2%; Score 19.8; DB 4; Length 634;
Best Local Similarity 91.3%; Pred. NO. 3.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAAGAAATTCAGTTCATAGCTTG 24
|||||
Db 29 GAAGAAATTCAGTTCCTAGCATG 51

RESULT 11
CD896013 636 bp mRNA linear EST 14-JUL-2003
LOCUS G174.101JL6R011120 G174 Triticum aestivum cDNA clone G174101J16,
DEFINITION mRNA sequence.
ACCESSION CD896013
VERSION CD896013.1 GI:32669222
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 636)
AUTHORS Genoplante.
TITLE Genoplante, a major partnership french program in plant genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).
FEATURES
source
1..636
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="G174101J16"
/tissue_type="grain (174 degrees per day after
pollination)"
/clone_lib="G174"
ORIGIN
Query Match 79.2%; Score 19.8; DB 6; Length 636;
Best Local Similarity 91.3%; Pred. NO. 3.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAAGAAATTCAGTTCATAGCTTG 24
|||||
Db 69 GAAGAAATTCAGTTCCTAGCATG 91

RESULT 12
CD453897 643 bp mRNA linear EST 03-JUN-2003
LOCUS WHE0902_E02_I04ZT CS wheat 5-15 DAP spike cDNA library Triticum
DEFINITION aestivum cDNA clone WHE0902_E02_I04, mRNA sequence.
ACCESSION CD453897
VERSION CD453897.1 GI:31368525
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.

```

REFERENCE 1 (bases 1 to 643)
 AUTHORS Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J. and Woo,J.
 TITLE The structure and function of the expressed portion of the wheat
 JOURNAL Unpublished (2003)
 COMMENT genomes - 5-15 DAP CS spike cDNA library
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oandern@pw.usda.gov
 This EST was generated by sequencing from the 3' end of the clone.
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20.
 Seq primer: T7 primer.

FEATURES
 source
 1..643
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE0902.E02_I04"
 /tissue_type="Spike"
 /dev_stage="Adult plant"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
 greenhouse. Spikes at 5, 10 and 15 DAP were harvested,
 total RNA and poly(A) RNA were prepared, a cDNA library
 was made, and the cDNA clones were in vivo excised to give
 phagescript phagemids in the TJ Close lab (Choi, Close,
 Fenton) at the University of California, Riverside.
 Plasmid DNA preparations and DNA sequencing were performed
 in the OD Anderson lab (all other authors)."

ORIGIN
 Query Match 79.2%; Score 19.8; DB 6; Length 643;
 Best Local Similarity 91.3%; Pred. No. 3.6e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24
 |||||
 DB 96 GAAGAAATTCAGTTCATAGCATG 118
 |||||

RESULT 13
 BQ804516/c
 LOCUS
 DEFINITION WHE3555_F02_K03ZS wheat developing grains cDNA library Triticum
 aestivum cDNA clone WHE3555_F02_K03, mRNA sequence.
 ACCESSION BQ804516
 VERSION BQ804516.1 GI:22028647
 EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 1 (bases 1 to 773)
 Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J.,
 Cronin,K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J.,
 Rausch,C.J., Wilson,C. and Woo,J.
 The structure and function of the expressed portion of the wheat
 genomes - Developing grains cDNA library
 Unpublished (2002)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773
 Fax: 5105595818
 Email: oandern@pw.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: SK primer.

FEATURES
 source
 1..773
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Butte 86"
 /db_xref="taxon:4565"
 /clone="WHE3555_F02_K03"
 /tissue_type="whole grains"
 /dev_stage="3-44 days post anthesis seed"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda ZAP II, excised phagemid; Site 1:
 EcoRI; Plants were grown under six following different
 environmental regimes in greenhouse, Environment 1)
 240C/170C day/night, well-watered, with post-anthesis
 fertilizer, Environment 2) 240C/170C day/night,
 well-watered, without post-anthesis fertilizer,
 Environment 3) 370C/170C day/night, well-watered, with
 post-anthesis fertilizer, Environment 4) 370C/170C
 day/night, well-watered, without post-anthesis fertilizer,
 Environment 5) 370C/170C day/night plus drought, with
 post-anthesis fertilizer, Environment 6) 370C/170C
 day/night plus drought, without post-anthesis fertilizer,
 and frozen in liquid nitrogen, Environment 1 at 3, 5, 7,
 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment
 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44
 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28,
 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20,
 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12,
 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10,
 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by
 S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A
 cDNA library was made using poly (A) RNA, and the cDNA
 clones were in vivo excised to give phagescript SK(-)
 phagemids in the TJ Close lab (Chin, Close, Fenton) at
 the University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (others)."

ORIGIN
 Query Match 79.2%; Score 19.8; DB 5; Length 773;
 Best Local Similarity 91.3%; Pred. No. 3.6e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAAGAAATTCAGTTCATAGCTTG 24
 |||||
 DB 754 GAAGAAATTCAGTTCCTAGCATG 732
 |||||

RESULT 14
 BZ186617
 LOCUS
 DEFINITION CH230-340J18_TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
 CH230-340J18, genomic survey sequence.
 ACCESSION BZ186617
 VERSION BZ186617.1 GI:23838548
 KEYWORDS GSS.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 783)
 Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
 Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D.,
 Riggs,F., de Jong,P. and Fraser,C.M.

TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other GSSs: CH230-340J18.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/or ering information.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
 Plate: 340 row: J column: 18
 Seq primer: T7
 Class: BAC ends.

FEATURES

source
 1. .783
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SnHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-340J18"
 /sex="Female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 2"
 /notes="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;
 CHORI-230 rat (BN/SnHsd/MCW) BAC library produced by
 Pieter de Jong"

ORIGIN

Query Match 79.2%; Score 19.8; DB 8; Length 783;
 Best Local Similarity 91.3%; Pred. No. 3.6e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCTT 23
 |||||
 Db 405 TGGAGAAATTCAGTTCATAGCTT 427

RESULT 15
CC495993
LOCUS CH240_331H6.T7 CHORI-240 Bos taurus genomic clone CH240_331H6,
DEFINITION genomic survey sequence.

ACCESSION CC495993
VERSION CC495993.1 GI:31809966
KEYWORDS GSS.

SOURCE

Bos taurus (cow)
 Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Rutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.

REFERENCE

1. (bases 1 to 863)
 Holt, R., Stott, J., Yang, G., Barber, S., Smailus, D., Prabhu, A.-L.,
 Tsai, M., Cloutier, A., Lee, D., Gilm, N., Olson, T., Mayo, M.,
 Butterfield, Y., Kirkpatrick, R., Liu, J., Guin, R., Chan, A., Chiu, R.,
 Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S.,
 Schein, J., Marra, M., de Jong, P., McWilliam, S., Barris, W.,
 Dalrymple, B. P. and Tellam, R.

TITLE Bovine BAC End Sequences from Library CHORI-240, PLATES 294 to 398
JOURNAL Unpublished (2003)
COMMENT Other GSSs: CH240_331H6.TARBAC13P2

Contact: Rob Holt
 Sequencing
 The British Columbia Cancer Agency Genome Science Centre
 600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
 Tel: 604-877-6085
 Fax: 604-877-6276
 Email: rholt@bcgsc.ca

Clones are derived from the bovine BAC library CHORI-240
 (http://www.chori.org/bacpac/bovine240.htm). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/ordering information.htm). This work
 was undertaken as part of the International Bovine BAC Mapping
 Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the
 British Columbia Genome Sciences Centre, Canada.
 Plate: 331 row: H column: 6
 Seq primer: T7
 Class: BAC ends.

FEATURES

source
 1. .863
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /strain="breed: Hereford"
 /db_xref="taxon:9913"
 /clone="CH240_331H6"
 /sex="Male"
 /cell_type="Blood"
 /clone_lib="CHORI-240"
 /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;
 Hereford bull Li Domino 99375; CHORI-240 Bovine BAC
 library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 79.2%; Score 19.8; DB 9; Length 863;
 Best Local Similarity 91.3%; Pred. No. 3.7e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAAGAAATTCAGTTCATAGCTT 23
 |||||
 Db 394 TGAAGAAATTCAGTTCATAGATT 416

Search completed: August 13, 2005, 06:45:05
 Job time: 1614.65 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 01:08:33 ; Search time 725.234 Seconds
(without alignments)
1536.704 Million cell updates/sec

Title: US-10-673-854-4

Perfect score: 23
Sequence: 1 caggagatcctgagattatgtgg 23

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ets:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| C 1 | 23 | 100.0 | 212280 | 9 HSB362E11 | AL121873 Human DNA |
| C 2 | 22 | 95.7 | 142099 | 9 AC002509 | AC002509 Homo sapi |
| C 3 | 22 | 95.7 | 176872 | 9 AC012078 | AC012078 Homo sapi |
| C 4 | 20.4 | 88.7 | 39904 | 9 AL590003 | AL590003 Human DNA |
| C 5 | 20.4 | 88.7 | 41502 | 9 AL513262 | AL513262 Human DNA |
| C 6 | 20.4 | 88.7 | 70140 | 9 AL139820 | AL139820 Human DNA |
| C 7 | 20.4 | 88.7 | 85000 | 9 HSA312686 | AJ312686 Homo sapi |
| C 8 | 20.4 | 88.7 | 86719 | 9 HSJ906P16 | AL079339 Human DNA |
| C 9 | 20.4 | 88.7 | 87817 | 9 AC079174 | AC079174 Homo sapi |
| C 10 | 20.4 | 88.7 | 88665 | 2 HSA132411 | AJ132411 Homo sapi |
| C 11 | 20.4 | 88.7 | 111026 | 2 AL157403 | AL157403 Homo sapi |
| C 12 | 20.4 | 88.7 | 113956 | 9 HS74M1 | AL035704 Human DNA |
| C 13 | 20.4 | 88.7 | 119012 | 9 AC106732 | AC106732 Homo sapi |
| C 14 | 20.4 | 88.7 | 128915 | 9 AC002416 | AC002416 Human Chr |
| C 15 | 20.4 | 88.7 | 139203 | 2 AC068311 | AC068311 Homo sapi |
| C 16 | 20.4 | 88.7 | 152531 | 9 AP005855 | AP005855 Homo sapi |
| C 17 | 20.4 | 88.7 | 156361 | 2 AC026174 | AC026174 Homo sapi |
| C 18 | 20.4 | 88.7 | 158028 | 2 AC019001 | AC019001 Homo sapi |
| C 19 | 20.4 | 88.7 | 159593 | 2 AC026302 | AC026302 Homo sapi |

| | | | | | |
|------|------|------|--------|--------------|--------------------|
| C 20 | 20.4 | 88.7 | 160373 | 2 AC024548 | AC024548 Homo sapi |
| C 21 | 20.4 | 88.7 | 162616 | 9 AC099539 | AC099539 Homo sapi |
| C 22 | 20.4 | 88.7 | 165261 | 2 AC012659 | AC012659 Homo sapi |
| C 23 | 20.4 | 88.7 | 175448 | 2 AC018651 | AC018651 Homo sapi |
| C 24 | 20.4 | 88.7 | 175598 | 2 AC136198 | AC136198 Papio anu |
| C 25 | 20.4 | 88.7 | 177253 | 9 AC092024 | AC092024 Homo sapi |
| C 26 | 20.4 | 88.7 | 181877 | 2 AC149458 | AC149458 Papio anu |
| C 27 | 20.4 | 88.7 | 193589 | 9 AC025370 | AC025370 Homo sapi |
| C 28 | 20.4 | 88.7 | 196361 | 9 AC005386 | AC005386 citb 57_1 |
| C 29 | 20.4 | 88.7 | 200034 | 9 AC107620 | AC107620 Homo sapi |
| C 30 | 20.4 | 88.7 | 203407 | 2 AC006174 | AC006174 Homo sapi |
| C 31 | 20.4 | 88.7 | 204116 | 9 CNS0000A | AL049828 Human chr |
| C 32 | 20.4 | 88.7 | 208547 | 2 AC021570 | AC021570 Homo sapi |
| C 33 | 20.4 | 88.7 | 347253 | 9 AF363578 | AF363578 Homo sapi |
| C 34 | 19.8 | 86.1 | 22955 | 6 AX780888 | AX780888 Sequence |
| C 35 | 19.8 | 86.1 | 22955 | 9 AK025451 | AK025451 Homo sapi |
| C 36 | 19.8 | 86.1 | 65557 | 9 AY485310 | AY485310 Homo sapi |
| C 37 | 19.8 | 86.1 | 80515 | 9 AC011992 | AC011992 Homo sapi |
| C 38 | 19.8 | 86.1 | 93288 | 9 AC020926 | AC020926 Homo sapi |
| C 39 | 19.8 | 86.1 | 105788 | 9 AC010458 | AC010458 Homo sapi |
| C 40 | 19.8 | 86.1 | 110000 | 2 AC145312_2 | Continuation (3 of |
| C 41 | 19.8 | 86.1 | 110000 | 2 AC024562_0 | Continuation (3 of |
| C 42 | 19.8 | 86.1 | 110000 | 2 AC024562_1 | Continuation (2 of |
| C 43 | 19.8 | 86.1 | 110000 | 2 AC024562_2 | Continuation (3 of |
| C 44 | 19.8 | 86.1 | 117213 | 9 AC026779 | AC026779 Homo sapi |
| C 45 | 19.8 | 86.1 | 132150 | 9 AC005586 | AC005586 Homo sapi |

ALIGNMENTS

RESULT 1
LOCUS HSB362E11/c 212280 bp DNA linear PRI 26-MAY-2000
DEFINITION Human DNA sequence from clone RP13-362E11 on chromosome X. Contains a pseudogene similar to mouse GEG-154 and moquitto MRRG, a pseudogene similar to human MMS2 and chicken CROC-1B, ESTs, STSs and GSSs, complete sequence.

ACCESSION AL121873
VERSION AL121873.15 GI:8218071
KEYWORDS HTG; CROC-1B; GEG-154; MMS2; MRRG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 212280)
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Wilson,S.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

COMMENT On Jun 3, 2000 this sequence version replaced gi:7329905.
During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX

RP13-362E11 is from the library RPCI-13.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
 VECTOR: pBACE3.6
 This sequence is the entire insert of clone RP13-362E11.

FEATURES

```

    source
      1..212280
        /organism="Homo sapiens"
        /mol_type="Genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="X"
        /clone="RP13-362E11"
        /clone_lib="RPCI-13.2"
    repeat_region
      5..42
        /notes="MER34 repeat: matches 498..534 of consensus"
    repeat_region
      120..403
        /notes="MER39b repeat: matches 68..453 of consensus"
    repeat_region
      404..498
        /notes="MER39 repeat: matches 7..103 of consensus"
    repeat_region
      502..1573
        /notes="L1M4 repeat: matches 4008..5095 of consensus"
    repeat_region
      1569..1914
        /notes="L1M4 repeat: matches 3525..3878 of consensus"
    repeat_region
      1906..2066
        /notes="L1MD repeat: matches 1376..1528 of consensus"
    repeat_region
      2112..2378
        /notes="L1MD repeat: matches 5..274 of consensus"
    repeat_region
      2379..2635
        /notes="L1MD repeat: matches 1126..1402 of consensus"
    repeat_region
      2651..2697
        /notes="L1MD repeat: matches 504..550 of consensus"
    repeat_region
      2840..3252
        /notes="L1MD repeat: matches 77..514 of consensus"
    repeat_region
      3351..3420
        /notes="L1MD repeat: matches 1..71 of consensus"
    repeat_region
      3421..3902
        /notes="HERVH repeat: matches 5231..5757 of consensus"
    repeat_region
      3903..4462
        /notes="280 copies 2 mer tt 57% conserved"
    repeat_region
      4465..4782
        /notes="L1UJo repeat: matches 1..312 of consensus"
    repeat_region
      4783..6107
        /notes="HERVL repeat: matches 3896..5242 of consensus"
    repeat_region
      6238..6378
        /notes="L1MD repeat: matches 973..1115 of consensus"
    repeat_region
      6640..6783
        /notes="L1MCS repeat: matches 7751..7910 of consensus"
    repeat_region
      7282..7567
        /notes="L1T1H repeat: matches 1..302 of consensus"
    misc_feature
      complement(7766..8138)
    misc_feature
      complement(7769..8123)
    misc_feature
      complement(7779..8205)
    repeat_region
      8002..8469
        /notes="L1T1D repeat: matches 1..503 of consensus"
    repeat_region
      8477..8654
        /notes="MIR repeat: matches 75..256 of consensus"
    repeat_region
      8715..8812
        /notes="49 copies 2 mer aa 61% conserved"
    gene
      complement(9391..9893)
        /gene="b3362E11.1"
        /pseudo
    CDS
      complement(9391..9893)
        /gene="b3362E11.1"
        /note="b3362E11.1 (pseudogene similar to mouse GEG-154 and mosquito MRRG)
        match: cDNAs: Em:AK001774
        match: proteins: Sw:P50636 Tr:Q9Y4Y1 Tr:P90662 Tr:P91067"
        /pseudo
        /codon_start=1
        /evidence=not_experimental

```

```

    misc_feature
      9919..10251
        /note="match: STS: Em:TI5642"
    repeat_region
      10912..13011
        /note="I1 repeat: matches 2373..4514 of consensus"
    repeat_region
      13018..13758
        /note="TIGGER1 repeat: matches 1..774 of consensus"
    repeat_region
      13759..13814
        /note="L1M4c repeat: matches 1557..1609 of consensus"
    repeat_region
      13815..14766
        /note="L1R5 repeat: matches 1..969 of consensus"
    repeat_region
      14767..14882
        /note="L1M4c repeat: matches 1609..1737 of consensus"
    repeat_region
      15152..16106
        /note="L1M3 repeat: matches 4670..5642 of consensus"
    repeat_region
      16107..16560
        /note="L1R46 repeat: matches 1..461 of consensus"
    repeat_region
      18749..19851
        /note="HERVH repeat: matches 3994..5500 of consensus"
    repeat_region
      19925..22648
        /note="HERVH21 repeat: matches 1275..3767 of consensus"
    misc_feature
      complement(22988..23428)
    misc_feature
      23057..23492
        /note="match: GSS: Em:AQ619480"
    repeat_region
      23749..24192
        /note="match: GSS: Em:AQ833119"
    repeat_region
      24194..24348
        /note="L1R46 repeat: matches 1..461 of consensus"
    repeat_region
      24386..24485
        /note="L1M3 repeat: matches 5639..5793 of consensus"
    repeat_region
      24497..25002
        /note="L1MA9 repeat: matches 5995..6096 of consensus"
    misc_feature
      complement(25454..26350)
    repeat_region
      25505..25526
        /note="match: GSS: Em:AQ894184"
    repeat_region
      25587..25823
        /note="I1 copies 2 mer ta 100% conserved"
    misc_feature
      complement(25882..26240)
    misc_feature
      25885..26180
        /note="match: GSS: Em:AQ754783"
    misc_feature
      complement(25910..26219)
    misc_feature
      25945..26256
        /note="match: GSS: Em:AQ785533"
    misc_feature
      25949..26676
        /note="match: GSS: Em:AQ628742"
    misc_feature
      25950..26147
        /note="match: GSS: Em:AL108886"
    misc_feature
      25951..26260
        /note="match: GSS: Em:AQ230662"
    misc_feature
      complement(25951..26227)
    misc_feature
      complement(25951..26105)
    repeat_region
      25951..25978
        /note="14 copies 2 mer ta 92% conserved"
    misc_feature
      25954..26352
        /note="match: GSS: Em:AQ525540"
    misc_feature
      complement(25978..26215)
    misc_feature
      complement(25979..26131)
    repeat_region
      25980..26025
        /note="23 copies 2 mer ta 84% conserved"
    misc_feature
      complement(25984..26583)
    misc_feature
      complement(25988..26185)
    misc_feature
      25994..26263
        /note="match: GSS: Em:AQ226794"
    misc_feature
      complement(25998..26169)

```


misc_feature /note="match: STS: Em:AU047022"
 complement(26000..26305)
 /note="match: GSS: Em:AQ123830"
 26005..26229
 /note="match: GSS: Em:AZ007683"
 26009..26160
 /note="match: STS: Em:HSB017WB9"
 26009..26147
 /note="match: GSS: Em:AZ069301"
 26037..26181
 /note="match: GSS: Em:AQ871738"
 complement(26046..26369)
 /note="match: GSS: Em:AQ827142"
 26046..26360

Query Match 100.0%; Score 23; DB 9; Length 212280;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGGAGATCTGAGATTATGTGG 23
 |||||
 Db 116939 CAGGAGATCTGAGATTATGTGG 116917

RESULT 2
 AC002509/c 142099 bp DNA linear PRI 25-NOV-1998
 LOCUS Homo sapiens chromosome Y, clone 2Y, complete sequence.
 DEFINITION AC002509
 ACCESSION AC002509.1 GI:3927860
 VERSION HTG.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 142099)
 Birren, B., Linton, J., Nussbaum, C. and Lander, E.
 Homo sapiens chromosome Y, clone 2Y
 Unpublished
 2 (bases 1 to 142099)
 Hawkins, T.L., Birren, B.W., Pasman, K.H., Nussbaum, C., Lander, E.S.,
 McKernan, K., Munro, C., Richardson, P., Barna, N., Chang, A., Cooke, P.,
 Daly, M.J., Devon, K., Dewar, K., Forrest, C., Gage, D., Geraghty, K.,
 Hagos, B., Huang, J., Hui, L., Jacotot, L., Kirby, A., Lane, M.,
 MacKenzie, J., Marquis, N., McDermott, J., Molla, M., Morrow, J.,
 Nachman, A., Naylor, J., Nussbaum, C., O'Connor, T., Olotu, A.,
 Peterson, K., Reeve, M.P., Roberts, D., Rollins, G., Stilwell, J.,
 Stone, C., Strickland, C., Sydney, K., Tang, L., Wilmer, F., Zemtseva, I.
 and Zody, M.

Direct Submission
 Submitted (27-AUG-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 142099)
 Birren, B., Linton, J., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baker, J., Baldwin, J., Barna, N., Beckler, R., Benn, J., Boutwell, C.,
 Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Corliss, D., Depayre, E., Devon, K., Dewar, K.,
 Donelan, L., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R.,
 Gage, D., Gardyna, S., Geraghty, K., Grant, G., Hagos, B., Headford, A.,
 Herena, L., Horton, L., Howland, J.C., Jacotot, L., Jones, C., Kann, L.,
 Karatas, A., Lehoczy, J., MacDonald, P., Marquis, N., McEwan, P.,
 McGurk, A., McKernan, K., Meldrim, J., Molla, M., Morris, W., Morrow, J.,
 Mychalek, J., Nahf, R., Naylor, J., Niloff, M., O'Connor, T., Roy, A.,
 O'Donnell, P., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Roy, A.,
 Severy, P., Stange-Thomann, N., Stilwell, J., Stojanovic, N., Stone, C.,
 Subramanian, A., Tesfaye, S., Tichovolsky, N., Torruella-Miller, I.,
 Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Wyman, D.,
 Ye, W.J., Zhao, J. and Zody, M.

Direct Submission
 Submitted (27-AUG-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 142099)

Birren, B., Linton, J., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baker, J., Baldwin, J., Barna, N., Beckler, R., Benn, J., Boutwell, C.,
 Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Corliss, D., Depayre, E., Devon, K., Dewar, K.,
 Donelan, L., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R.,
 Gage, D., Gardyna, S., Geraghty, K., Grant, G., Hagos, B., Headford, A.,
 Herena, L., Horton, L., Howland, J.C., Jacotot, L., Jones, C., Kann, L.,
 Karatas, A., Lehoczy, J., MacDonald, P., Marquis, N., McEwan, P.,
 McGurk, A., McKernan, K., Meldrim, J., Molla, M., Morris, W., Morrow, J.,
 Mychalek, J., Nahf, R., Naylor, J., Niloff, M., O'Connor, T., Roy, A.,
 O'Donnell, P., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Roy, A.,
 Severy, P., Stange-Thomann, N., Stilwell, J., Stojanovic, N., Stone, C.,
 Subramanian, A., Tesfaye, S., Tichovolsky, N., Torruella-Miller, I.,
 Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Wyman, D.,
 Ye, W.J., Zhao, J. and Zody, M.

Direct Submission
 Submitted (25-NOV-1998) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Nov 25, 1998 this sequence version replaced gi:3924665.
 All repeats were identified using RepeatMasker: Smit, A.F.A. &

Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.
 Location/Qualifiers
 1..142099
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="Y"
 /map="Y"
 /clone="2Y"
 /clone_lib="unknown"
 complement(315..438)
 /rpt_family="MER61-internal"
 complement(2160..2521)
 /rpt_family="MER50"
 2522..2578
 /rpt_family="AluSg/x"
 2859..2919
 /rpt_family="MERS1-internal"
 complement(2920..2978)
 /rpt_family="(CA)n"
 2990..3699
 /rpt_family="MER4-internal"
 complement(3700..3814)
 /rpt_family="LTR8"
 3815..4468
 /rpt_family="MER4-internal"
 4523..7011
 /rpt_family="MER4-internal"
 complement(7295..7598)
 /rpt_family="AluSp"
 9162..9349
 /rpt_family="MER4-internal"
 9361..9509
 /rpt_family="MER4A2"
 9510..9811
 /rpt_family="MER4A"
 complement(9812..9962)
 /rpt_family="LIM1"
 complement(10212..10830)
 /rpt_family="LIM4"
 complement(10831..11120)
 /rpt_family="AluY"
 complement(11121..11337)
 /rpt_family="LIM4"
 complement(11508..12117)
 /rpt_family="LIMBc"
 complement(12315..12626)
 /rpt_family="LIMB"
 complement(12923..13136)
 /rpt_family="MIR"
 complement(13246..13328)
 /note="Single-stranded terminator coverage."
 13761..13782
 /rpt_family="AT_rich"
 14930..14950
 /rpt_family="AT_rich"
 15128..15231
 /rpt_family="HAL1"
 15292..15552
 /rpt_family="HAL1"
 complement(15690..16275)
 /rpt_family="MER4B"
 complement(16397..16551)
 /rpt_family="MER66-internal"
 16825..16846
 /rpt_family="AT_rich"
 complement(17272..17680)
 /rpt_family="MER4-internal"
 complement(17687..17920)
 /rpt_family="MER31-internal"
 complement(18046..18789)
 /rpt_family="MER50"

```
repeat_region complement(19887..20110)
repeat_region /rpt_family="MER57-internal"
repeat_region complement(20133..20299)
repeat_region /rpt_family="MER4-internal"
repeat_region 20300..20593
repeat_region /rpt_family="AluY"
repeat_region complement(20594..20673)
repeat_region /rpt_family="MER4-internal"
repeat_region complement(20674..20974)
repeat_region /rpt_family="AluSg"
repeat_region complement(20975..21867)
repeat_region /rpt_family="MER4-internal"
repeat_region complement(21372..21498)
repeat_region /rpt_family="MER51-internal"
repeat_region complement(21499..21804)
repeat_region /rpt_family="AluSg"
repeat_region complement(21805..21964)
repeat_region /rpt_family="MER51-internal"
repeat_region complement(21965..22251)
repeat_region /rpt_family="MER4-internal"
repeat_region complement(22252..22441)
repeat_region /rpt_family="MER51-internal"
repeat_region complement(22490..22671)
repeat_region /rpt_family="MER83-internal"
repeat_region complement(22589..22693)
repeat_region /rpt_family="MER65-internal"
repeat_region 23324..23353
repeat_region /rpt_family="AT_rich"
repeat_region complement(23437..23977)
repeat_region /rpt_family="MER4B"
repeat_region complement(23978..24177)
repeat_region /rpt_family="MLT2A"
repeat_region 24557..24891
repeat_region /rpt_family="HAL1"
repeat_region 25476..25758
repeat_region /rpt_family="L2"
repeat_region 26069..26101
repeat_region /rpt_family="AT_rich"
repeat_region 26265..26489
repeat_region /rpt_family="MLT1D"
repeat_region 26490..26517
repeat_region /rpt_family="L2"
repeat_region 26518..26780
repeat_region /rpt_family="MLT1D"
repeat_region 27156..27180
repeat_region /rpt_family="AT_rich"
repeat_region complement(27417..27797)
repeat_region /rpt_family="L2"
repeat_region complement(27963..28332)
repeat_region /rpt_family="MSTA"
repeat_region complement(28333..29268)
repeat_region /rpt_family="MSTA-internal"
repeat_region 29269..29580
repeat_region /rpt_family="AluY"
repeat_region complement(29581..30301)
repeat_region /rpt_family="MSTA-internal"
repeat_region complement(30304..30679)
repeat_region /rpt_family="MSTA"
repeat_region complement(30680..31159)
repeat_region /rpt_family="LTR40a"
repeat_region complement(31325..31655)
repeat_region /rpt_family="LINE"
repeat_region 31692..31731
repeat_region /rpt_family="AT_rich"
repeat_region 31883..31927
repeat_region /rpt_family="AT_rich"
repeat_region complement(32107..33826)
repeat_region /rpt_family="LIM4"
repeat_region 33831..34558
repeat_region /rpt_family="LINE3"
repeat_region 34987..35012
repeat_region /rpt_family="AT_rich"
repeat_region 35108..35669

/rpt_family="MER4B"
36280..36322
/rpt_family="AT_rich"
complement(36386..36519)
/rpt_family="MIR"
complement(36755..37203)
/rpt_family="MER4C"

Query Match 95.7%; Score 22; DB 9; Length 142099;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 35610 CAGGAGATCCTGAGATTATGTG 35589

RESULT 3
AC012078/c 176872 bp DNA linear PRI 30-SEP-2000
LOCUS Homo sapiens BAC clone RP11-539022 from Y, complete sequence.
DEFINITION AC012078
ACCESSION AC012078
VERSION AC012078.3 GI:7684580
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 176872)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 176872)
AUTHORS Hou,S., Maupin,R. and Gibson,A.
TITLE The sequence of Homo sapiens BAC clone RP11-539022
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 176872)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 176872)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 176872)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 176872)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 3, 2000 this sequence version replaced gi:7631053.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_NH0539022
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
```

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Ooegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-530K5; the clone sequenced to the right is RP11-33605. Actual start of this clone is at base position 1 of RP11-539022; actual end is at base position 176872 of RP11-539022.

FEATURES

```

Source
1. 176872
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="Y"
   /map="Y"
   /clone="RP11-539022"
   /clone_lib="RPCI-11"
repeat_region 1. 204
   /rpt_family="L1"
repeat_region 202. 1293
   /rpt_family="L1"
repeat_region 1294. 1324
   /rpt_family="L1"
repeat_region 1325. 2158
   /rpt_family="L1"
repeat_region 2159. 2339
   /rpt_family="L1"
repeat_region 2340. 3099
   /rpt_family="L1"
repeat_region 3120. 3219
   /rpt_family="Alu"
repeat_region 3220. 3945
   /rpt_family="L1"
repeat_region 3945. 4535
   /rpt_family="L1"
repeat_region 4565. 4619
   /rpt_family="Alu"
repeat_region 4638. 5206
   /rpt_family="L1"
repeat_region 5255. 5268
   /rpt_family="Alu"
repeat_region 5269. 5291
   /rpt_family="(TATG)n"
repeat_region 5292. 5583
   /rpt_family="Alu"
repeat_region 5989. 6021
   /rpt_family="AT_rich"

```

```

repeat_region 6248. 6268
   /rpt_family="AT_rich"
repeat_region 6600. 6686
   /rpt_family="MIR"
repeat_region 7754. 8717
   /rpt_family="L1"
repeat_region 8711. 10078
   /rpt_family="L1"
repeat_region 10079. 10455
   /rpt_family="MaLR"
repeat_region 10456. 11631
   /rpt_family="L1"
repeat_region 11657. 11760
   /rpt_family="L1"
repeat_region 11761. 12278
   /rpt_family="Retroviral"
repeat_region 12282. 12941
   /rpt_family="L1"
misc_feature 12940. 12962
   /note="similar to EST A1095060 (NID:G3434036) gal0c03.xl"
repeat_region 12942. 13243
   /rpt_family="Alu"
repeat_region 13241. 14024
   /rpt_family="L1"
repeat_region 14025. 14071
   /rpt_family="L1"
repeat_region 14072. 14107
   /rpt_family="(TAA)n"
repeat_region 14108. 14764
   /rpt_family="L1"
repeat_region 14770. 14800
   /rpt_family="AT_rich"
repeat_region 15566. 15598
   /rpt_family="AT_rich"
repeat_region 15847. 15867
   /rpt_family="AT_rich"
repeat_region 16123. 16431
   /rpt_family="Alu"
repeat_region 17318. 17558
   /rpt_family="L1"
repeat_region 17655. 18325
   /rpt_family="L1"
repeat_region 18319. 18728
   /rpt_family="L1"
repeat_region 18735. 18866
   /rpt_family="L1"
repeat_region 18880. 19147
   /rpt_family="MaLR"
repeat_region 19173. 20018
   /rpt_family="L1"
repeat_region 20015. 20440
   /rpt_family="MER4-group"
repeat_region 20441. 20536
   /rpt_family="LTR54"
repeat_region 20537. 21186
   /rpt_family="MER4-group"
repeat_region 21195. 21580
   /rpt_family="L1"
repeat_region 21581. 21618
   /rpt_family="(TTTA)n"
repeat_region 21671. 21838
   /rpt_family="L1"
repeat_region 21839. 22135
   /rpt_family="Alu"
repeat_region 22136. 22315
   /rpt_family="L1"
repeat_region 24191. 24210
   /rpt_family="(TTG)n"
repeat_region 24211. 24482
   /rpt_family="Alu"
repeat_region 24828. 25925
   /rpt_family="MER4-group"
repeat_region 25914. 26164

```

Query Match 95.7%; Score 22; DB 9; Length 176872;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGAGATCTGAGATTATGTG 22

Db 170703 CAGGAGATCTGAGATTATGTG 170682

RESULT 4
 AL590003 39904 bp DNA linear PRI 06-APR-2001
 LOCUS Human DNA sequence from clone RP1-235I20 on chromosome 6, complete
 DEFINITION sequence.
 ACCESSION AL590003
 VERSION AL590003.4 GI:13560040
 KEYWORDS HTG.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 39904)
 Phillimore.B.

REFERENCE Direct Submission
 Submitted (05-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk

On Apr 6, 2001 this sequence version replaced gi:13446547.
 During cloning assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em., EMBL; SW,
 SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>

RP1-235I20 is from the library RPCI-1 constructed by the group of
 Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pCVPAC2
 IMPORTANT: This sequence is not the entire insert of clone
 RP1-235I20 It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP1-235I20 is at 1 in this sequence. The
 true left end of clone RP11-559K11 is at 39805 in this sequence.

FEATURES

source

1. .39904
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosomes="6"
 /clone="RP1-235I20"
 /clone_lib="RPCI-1"
 /notes="L2 repeat: matches 2518. .2750 of consensus"
 3626. .4121
 /notes="match: GSS: Em:AQ404463"
 4234. .4378

repeat_region

1724. .1957

misc_feature

3626. .4121

repeat_region

4234. .4378

/note="L1LWAS repeat: matches 6141. .6287 of consensus"
 4373. .4627
 /note="L1LWAS repeat: matches 5903. .6156 of consensus"
 5993. .6167
 /note="MER5B repeat: matches 1. .175 of consensus"
 6358. .6756
 /note="match: GSS: Em:AQ404373"
 6369. .6934
 /note="match: GSS: Em:B88668
 match: STS: Em:G49881"
 7400. .7769
 /note="THE1C repeat: matches 2. .371 of consensus"
 8020. .8356
 /note="MLRID repeat: matches 3. .349 of consensus"
 8727. .8953
 /note="MLT1J repeat: matches 186. .420 of consensus"
 9112. .9143
 /note="16 copies 2 mer gt 87% conserved"
 9805. .10110
 /note="AluSg repeat: matches 1. .305 of consensus"
 complement (10204. .10572)
 /note="match: GSS: Em:AQ606186"
 11347. .11573
 /note="WIR repeat: matches 32. .261 of consensus"
 12274. .12753
 /note="match: GSS: Em:AQ356231"
 13792. .13854
 /note="3 copies 21 mer 84% conserved"
 14365. .14669
 /note="AluSx repeat: matches 5. .310 of consensus"
 14906. .15223
 /note="L1MC4 repeat: matches 7249. .7586 of consensus"
 15269. .15387
 /note="WIR repeat: matches 20. .140 of consensus"
 16585. .16863
 /note="AluWb repeat: matches 18. .290 of consensus"
 16871. .17108
 /note="AluY repeat: matches 73. .310 of consensus"
 17130. .17240
 /note="MER45B repeat: matches 922. .1039 of consensus"
 17538. .17567
 /note="15 copies 2 mer ac 100% conserved"
 18069. .18287
 /note="MLT1A1 repeat: matches 142. .365 of consensus"
 18321. .18855
 /note="MER4D repeat: matches 431. .973 of consensus"
 complement (18669. .19138)
 /note="match: GSS: Em:AQ403612"
 complement (18696. .19162)
 /note="match: GSS: Em:AQ617634"
 18895. .19009
 /note="MLT1A1 repeat: matches 1. .115 of consensus"
 19067. .19149
 /note="MER5B repeat: matches 60. .129 of consensus"
 19165. .19986
 /note="match: GSS: Em:AQ751839"
 19275. .19414
 /note="L2 repeat: matches 2602. .2736 of consensus"
 19909. .20111
 /note="L1PB1 repeat: matches 5944. .6154 of consensus"
 20199. .22989
 /note="L1PA8 repeat: matches 3384. .6163 of consensus"
 22990. .23303
 /note="AluSx repeat: matches 1. .307 of consensus"
 23304. .26961
 /note="L1PA8 repeat: matches 6. .3303 of consensus"
 27432. .27889
 /note="match: GSS: Em:AQ439411"
 28292. .28402
 /note="MER81 repeat: matches 2. .114 of consensus"
 28573. .28617
 /note="MER58 repeat: matches 2248. .2292 of consensus"
 29055. .29133

repeat_region /notes="LTR40a repeat: matches 62. .146 of consensus" 30498. .30597
repeat_region /notes="L2 repeat: matches 2629. .2732 of consensus" 31541. .31676
repeat_region /notes="MT1E repeat: matches 257. .392 of consensus" 31647. .31768
repeat_region /notes="Murid repeat: matches 326. .462 of consensus" 32760. .33055
repeat_region /notes="MT1J repeat: matches 214. .505 of consensus" 33110. .33413
repeat_region /notes="AluSg repeat: matches 1. .310 of consensus" 35009. .35548
misc_feature /notes="match: GSS: Em:AQ602626" 36014. .36132
repeat_region /notes="MER86 repeat: matches 19. .149 of consensus" 36608. .36657
repeat_region /notes="L1PB3 repeat: matches 6101. .6150 of consensus" 37258. .39904
repeat_region /notes="L1MAL repeat: matches 3664. .6297 of consensus"

ORIGIN

Query Match 88.7%; Score 20.4; DB 9; Length 39504;
Best Local Similarity 95.5%; Pred. No. 9.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 18372 CAGGAGATCCTGAGATTATGTG 18393

RESULT 5
AL513262
LOCUS Human DNA sequence from clone RP11-499115 on chromosome 1, complete sequence.
DEFINITION
ACCESSION AL513262.7 GI:17381349
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (30-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Dec 5, 2001 this sequence version replaced gi:16944082.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SW., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-499115 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-499115 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-54388 is at 39503 in this sequence. The true right end of clone RP4-553F17 is at 2000 in this sequence.

FEATURES
Location/Qualifiers
source 1. .41502
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-499115"
/clone_lib="RPC1-11.2"

ORIGIN

Query Match 88.7%; Score 20.4; DB 9; Length 41502;
Best Local Similarity 95.5%; Pred. No. 9.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 37869 CAGGAGATCCTGAGATTATGTG 37890

RESULT 6
AL139820
LOCUS Human DNA sequence from clone RP11-396A24 on chromosome 10, complete sequence.
DEFINITION
ACCESSION AL139820
VERSION AL139820.10 GI:14272240
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (30-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
JOURNAL requests: clonerequest@sanger.ac.uk
COMMENT On May 31, 2001 this sequence version replaced gi:13751278.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SW., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-396A24 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-396A24 It may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.
The true right end of clone RP11-396A24 is at 70140 in this sequence. The true left end of clone RP11-343D22 is at 5760 in this sequence. The true right end of clone RP11-168C9 is at 100 in this sequence.

FEATURES

Location/Qualifiers

| | | |
|---------------|--|--|
| source | 1. .70140 | |
| | /organism="Homo sapiens" | |
| | /mol_type="genomic DNA" | |
| | /db_xref="taxon:9606" | |
| | /chromosome="10" | |
| | /clone="RP11-396A24" | |
| | /clone_lib="RPC1-11.2" | |
| repeat_region | 1. .315 | /note="MLT2G repeat: matches 3. .268 of consensus" |
| repeat_region | 324. .1989 | /note="MLT2G repeat: matches 3. .268 of consensus" |
| repeat_region | 2022. .2050 | /note="HERVL repeat: matches 4026. .5750 of consensus" |
| misc_feature | complement(2274. .2851) | /note="HERVL repeat: matches 3445. .3473 of consensus" |
| misc_feature | 2847. .3318 | /note="match: GSS: Em:AQ263367" |
| | /note="match: GSS: Em:AQ567074" | |
| repeat_region | 2972. .3102 | /note="L2 repeat: matches 2572. .2750 of consensus" |
| repeat_region | 3473. .3683 | /note="L2 repeat: matches 40. .260 of consensus" |
| repeat_region | 3887. .3998 | /note="MIR repeat: matches 40. .260 of consensus" |
| repeat_region | 4502. .4659 | /note="MIR repeat: matches 59. .178 of consensus" |
| repeat_region | 4671. .4795 | /note="MER5A repeat: matches 2. .170 of consensus" |
| | /note="MER81 repeat: matches 1. .114 of consensus" | |
| misc_feature | 5762. .6007 | /note="match: GSS: Em:AQ395191" |
| repeat_region | 5793. .6215 | /note="MLT1A2 repeat: matches 1. .405 of consensus" |
| misc_feature | 6322. .6496 | /note="match: GSS: Em:AQ815465" |
| misc_feature | 6338. .6628 | /note="match: GSS: Em:AQ517696" |
| repeat_region | 6370. .6938 | /note="MER4B repeat: matches 1. .575 of consensus" |
| repeat_region | 8625. .8926 | /note="AluY repeat: matches 1. .302 of consensus" |
| misc_feature | complement(8707. .9267) | /note="match: GSS: Em:AQ307791" |
| repeat_region | 9958. .10246 | /note="AluSg repeat: matches 1. .289 of consensus" |
| misc_feature | complement(11298. .11808) | /note="match: STS: Em:G49507" |
| misc_feature | complement(11361. .11848) | /note="match: GSS: Em:AQ481123" |
| misc_feature | 11868. .12312 | /note="match: GSS: Em:AQ369493" |
| repeat_region | 13351. .13454 | /note="2 copies 52 mer 92% conserved" |
| repeat_region | 13514. .13823 | /note="AluJo repeat: matches 6. .310 of consensus" |
| misc_feature | 14996. .15576 | /note="match: STS: Em:G58481" |
| repeat_region | 18117. .18286 | /note="5 copies 34 mer 74% conserved" |
| repeat_region | 18122. .18277 | /note="3 copies 52 mer 78% conserved" |
| repeat_region | 18123. .18218 | /note="78 copies 2 mer ca 76% conserved" |
| misc_feature | complement(19440. .19834) | /note="match: GSS: Em:AQ12568" |
| repeat_region | 22796. .22835 | /note="20 copies 2 mer ct 87% conserved" |
| repeat_region | 22803. .22838 | |
| repeat_region | 23711. .23836 | /note="6 copies 6 mer tctctc 91% conserved" |
| repeat_region | 25064. .25517 | /note="63 copies 2 mer ta 58% conserved" |
| misc_feature | 27070. .27599 | /note="MLT1D repeat: matches 1. .463 of consensus" |
| | /note="match: GSS: Em:AQ601725" | |
| repeat_region | 27254. .27421 | /note="FRAM repeat: matches -1. .168 of consensus" |
| repeat_region | 27781. .28247 | /note="HAL1 repeat: matches 419. .886 of consensus" |
| repeat_region | 28623. .28757 | /note="MIR repeat: matches 106. .243 of consensus" |
| misc_feature | 28916. .29598 | /note="match: GSS: Em:AQ390359" |
| repeat_region | 29470. .30077 | /note="L1MB2 repeat: matches 5551. .6167 of consensus" |
| repeat_region | 30434. .30944 | /note="L1MB3 repeat: matches 5375. .5901 of consensus" |
| repeat_region | 31038. .31341 | /note="AluSx repeat: matches 1. .304 of consensus" |
| repeat_region | 31427. .31906 | /note="L1MBE repeat: matches 5226. .5728 of consensus" |
| repeat_region | 32282. .32686 | /note="L1M4 repeat: matches 3515. .3937 of consensus" |
| repeat_region | 32818. .33025 | /note="L1MC5 repeat: matches 7355. .7555 of consensus" |
| repeat_region | 33210. .33249 | /note="20 copies 2 mer tt 85% conserved" |
| misc_feature | complement(33442. .33831) | /note="match: GSS: Em:B52135" |
| misc_feature | 34349. .34584 | /note="match: GSS: Em:AQ416077" |
| repeat_region | 34611. .34920 | /note="MLT1D repeat: matches 191. .533 of consensus" |
| repeat_region | 36001. .36065 | /note="L1M3e repeat: matches 209. .282 of consensus" |
| repeat_region | 36300. .36400 | /note="L2 repeat: matches 2607. .2710 of consensus" |
| repeat_region | 37143. .37455 | /note="AluSx repeat: matches 1. .312 of consensus" |
| misc_feature | complement(37724. .38564) | /note="match: GSS: Em:AQ738968" |
| repeat_region | 37958. .38287 | /note="MLT1C repeat: matches 144. .466 of consensus" |
| repeat_region | 38288. .38576 | /note="AluJo repeat: matches 13. .297 of consensus" |
| misc_feature | 38570. .39096 | /note="match: GSS: Em:AQ337467" |
| repeat_region | 38577. .38682 | /note="MLT1C repeat: matches 38. .144 of consensus" |
| repeat_region | 41198. .41284 | /note="MIR repeat: matches 81. .173 of consensus" |
| repeat_region | 41532. .41874 | /note="MER47A repeat: matches 1. .366 of consensus" |
| repeat_region | 41878. .41975 | /note="MIR repeat: matches 59. .164 of consensus" |
| repeat_region | 42637. .42726 | /note="MIR repeat: matches 71. .164 of consensus" |
| misc_feature | complement(44297. .44694) | /note="match: GSS: Em:AQ545169" |
| repeat_region | 45790. .46084 | /note="AluSg repeat: matches 6. .297 of consensus" |
| repeat_region | 47341. .47673 | /note="MLT2E repeat: matches 1. .352 of consensus" |
| repeat_region | 47675. .47722 | /note="24 copies 2 mer ag 81% conserved" |
| repeat_region | 47687. .47722 | /note="6 copies 6 mer agagag 88% conserved" |
| misc_feature | complement(48625. .49183) | /note="match: GSS: Em:AQ556578" |
| misc_feature | 49192. .49631 | /note="match: GSS: Em:AQ675010" |

```

repeat_region 49940..50023
/notes="14 copies 6 mer ccat 70% conserved"
repeat_region 50023..50323
/notes="AluSp repeat: matches 2..303 of consensus"
repeat_region 52327..52616
/notes="AluX repeat: matches 1..300 of consensus"
repeat_region 54026..54434
/notes="MLT2FB repeat: matches 4..408 of consensus"
repeat_region 54456..54874

Query Match      88.7%; Score 20.4; DB 9; Length 70140;
Best Local Similarity 95.5%; Pred. No. 9.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 6426 CAGGAGATCCTGAGATTATGTG 6447

RESULT 7
HSA312686      85000 bp DNA linear PRI 16-MAY-2002
LOCUS
DEFINITION Homo sapiens partial LIMD1 gene for LIM domains containing 1, exons
1-2, complete sequence.
ACCESSION AJ312686
VERSION AJ312686.1 GI:13548631
KEYWORDS HTG; LIM domains containing 1; LIMD1 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kiss,H., Yang,X., Kiss,C., Andersson,K., Klein,G., Imreh,S. and
Dumanski,J.P.
TITLE The transcriptional map of the common eliminated region 1 (C3CER1)
in 3p21.3
JOURNAL Eur. J. Hum. Genet. 10 (1), 52-61 (2002)
MEDLINE 21906202
PUBMED 11896456
REFERENCE 2 (bases 1 to 85000)
AUTHORS Kiss,H.
TITLE Direct Submission
JOURNAL Submitted (01-APR-2001) Kiss H., Microbiology and Tumorbilogy
Center (MTC), Karolinska Institute, Box 280, Stockholm, S-17177,
SWEDEN
COMMENT On Apr 9, 2001 this sequence version replaced gi:6599071.
FEATURES
source 1..85000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p21.3"
/clone="RP5-1033n4"
24287..65692
/genes="LIMD1"
/notes="LIMD1"
24287..25715
/genes="LIMD1"
/number=1
Join(24308..25715,65591..65692)
/genes="LIMD1"
/codon_start=1
/product="LIM domains containing 1"
/protein_id="CAC35917.1"
/db_xref="GI:13548632"
/db_xref="GOA:Q9UGF4"
/db_xref="UniProt/TREMBL:Q9UGF4"
/translation="MDKYDDLGLAEASKFTIEDLMVYASDKGLFRVDKAGNNPEFET
RRVFATMAKHLQCCQQLQOETLPRGSRPVNGGRLGPOARWEVVGSKLTVDGA
AKPLAATGACPAVNTLAAGQPPYPQQRSPRLHGTGRHSQDCGSRSLATSEMS
AFHQPGCEDPCTGHDYDNLUSLASPKWGDKPGVPSIGLSVGSGWPSRSGSDPL
PKPGDHLNHLQLSRSSSESGSLGSGGNSSEKPTGLWSTASSQRVSPGLP
SPNLENGAPAVGVPQRTPTSVSAPLALSCPRQGLPRNSGLGSGVSGVMSKPNVDPL

```

```

PMFQDPKSYLSSAPSSPAGLDGSGQAVFGLGPKPGCTDLGTGPKLSPTSLVHPV
MSTLPELSCKEGLGWSGSLGSLVLDSPSPRVRLPCQPLVPGELRPSAELEKLE
AUTQRLREMDAHPKADYFGACVKCSKGVFGAGQACQAMGNLYHDTCTCAAC"
intron 25716..65590
/genes="LIMD1"
/number=1
exon 65591..65692
/genes="LIMD1"
/number=2

ORIGIN
Query Match      88.7%; Score 20.4; DB 9; Length 85000;
Best Local Similarity 95.5%; Pred. No. 9.7;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 17729 CAGGAGATCCTGAGATTATGTG 17750

RESULT 8
HSJ906P16      86719 bp DNA linear PRI 10-MAR-2001
LOCUS
DEFINITION Human DNA sequence from clone RP5-906P16 on chromosome
20q13.12-13.2 Contains ESTs, STSS and GSSs, complete sequence.
ACCESSION AL079339
VERSION AL079339.11 GI:6010222
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 86719)
AUTHORS Sehra,H.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Oct 4, 1999 this sequence version replaced gi:5918464.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was finished as follows unless otherwise noted: all regions were
either double-stranded or sequenced with an alternate chemistry or
covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at least one
plasmid subclone or more than one M13 subclone; and the assembly
was confirmed by restriction digest. This sequence was generated
from part of bacterial clone contigs of human chromosome 20,
constructed by the Sanger Centre Chromosome 20 Mapping Group.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP5-906P16 is from the library RPCI-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP5-906P16 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP5-906P16 is at 1 in this sequence. The
true left end of clone RP5-1009H6 is at 86620 in this sequence. The
true right end of clone RP5-1106N18 is at 58370 in this sequence.
Location/Qualifiers
1..86719
source

```


/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="q13.12-13.2"
/clone="RP5-906P16"
/clone_lib="RPC1-5"
415..881
/note="L1ME repeat: matches 5318..5783 of consensus"
1170..11730
/note="L1MB5 repeat: matches 5588..6184 of consensus"
1885..2135
/note="AluJo repeat: matches 2..278 of consensus"
2136..2187
/note="26 copies 2 mer ac 84% conserved"
2602..2763
/note="MIR repeat: matches 107..262 of consensus"
2779..3035
/note="L2 repeat: matches 2245..2481 of consensus"
3626..3716
/note="MIR repeat: matches 28..126 of consensus"
3891..4163
/note="MLT1B repeat: matches 1..246 of consensus"
4164..4488
/note="AluX repeat: matches 5..311 of consensus"
4489..4624
/note="MLT1B repeat: matches 246..388 of consensus"
complement(4971..5437)
/note="match: GSS: Em:AQ415388"
complement(5094..5437)
/note="match: GSS: Em:AQ402893"
complement(5119..5437)
/note="match: GSS: Em:AQ379417"
complement(5136..5605)
/note="match: GSS: Em:AQ135679"
5275..5352
/note="MIR repeat: matches 65..142 of consensus"
5438..5600
/note="MIR repeat: matches 91..262 of consensus"
5627..5921
/note="L2 repeat: matches 2398..2750 of consensus"
5922..6035
/note="MER58A repeat: matches 1..8 of consensus"
6036..6348
/note="AluS repeat: matches 1..309 of consensus"
6349..6481
/note="MER58A repeat: matches 8..223 of consensus"
6482..6522
/note="L2 repeat: matches 2362..2398 of consensus"
complement(6523..6665)
/note="match: GSS: Em:B44279"
6666..7162
/note="LTR19A repeat: matches 1..486 of consensus"
7627..7717
/note="MIR repeat: matches 97..196 of consensus"
7720..7935
/note="MER58 repeat: matches 1..2462 of consensus"
8075..8323
/note="MIR repeat: matches 2..260 of consensus"
8456..8766
/note="L1P4 repeat: matches 5836..6146 of consensus"
8767..8814
/note="24 copies 2 mer aa 83% conserved"
9106..9235
/note="MIR repeat: matches 9..138 of consensus"
9856..9940
/note="L2 repeat: matches 2624..2709 of consensus"
9954..10663
/note="LTR28 repeat: matches 300..1688 of consensus"
10634..10704
/note="LTR1 repeat: matches 119..190 of consensus"
10729..10893
/note="MER61B repeat: matches 1..106 of consensus"

repeat_region
10963..11243
/note="L2 repeat: matches 2217..2496 of consensus"
11391..11522
/note="MER4D repeat: matches 824..972 of consensus"
complement(11523..12124)
/note="match: GSS: Em:AQ265342"
11524..12038
/note="MER4A2 repeat: matches 16..502 of consensus"
complement(12039..12124)
/note="match: GSS: Em:AQ279178"
12125..12530
/note="MER4D repeat: matches 388..793 of consensus"
12531..12760
/note="MER4D repeat: matches 1..234 of consensus"
13775..14083
/note="AluSg repeat: matches 1..309 of consensus"
15616..15816
/note="MER58A repeat: matches 1..203 of consensus"
16071..16175
/note="LTR33 repeat: matches 100..214 of consensus"
16411..16604
/note="L1ME repeat: matches 5595..5786 of consensus"
16645..17075
/note="MER83 repeat: matches 1..448 of consensus"
17376..17461
/note="L1ME3A repeat: matches 6048..6145 of consensus"
17462..17768
/note="AluJo repeat: matches 1..302 of consensus"
17769..17785
/note="L1ME3A repeat: matches 6145..6159 of consensus"
17791..17932
/note="MSTA repeat: matches 1..160 of consensus"
17968..18576
/note="MER4B repeat: matches 1..569 of consensus"
18582..18629
/note="MSTA repeat: matches 172..217 of consensus"
18623..18688
/note="MSTB repeat: matches 360..426 of consensus"
18709..19070
/note="MER61B repeat: matches 3..396 of consensus"
19147..19374
/note="MER46A repeat: matches 1..236 of consensus"
19567..19655
/note="MIR repeat: matches 78..169 of consensus"
complement(19656..20051)
/note="match: GSS: Em:AQ587104"
complement(19725..20051)
/note="match: GSS: Em:AQ123463"
20052..20254
/note="MER20 repeat: matches 4..210 of consensus"
complement(20255..20679)
/note="match: GSS: Em:AQ382186"
complement(20301..20679)
/note="match: GSS: Em:AQ0709243"
complement(20341..20679)
/note="match: GSS: Em:AQ121387"
20680..20850
/note="MLT1E repeat: matches 291..463 of consensus"
21477..21878
/note="match: GSS: Em:AQ038484"
22066..22198
/note="MIR repeat: matches 12..156 of consensus"
22274..22572
/note="AluX repeat: matches 1..292 of consensus"
22588..22841
/note="L2 repeat: matches 2357..2608 of consensus"
23153..23449
/note="MIR1E repeat: matches 245..566 of consensus"
23485..23536
/note="26 copies 2 mer ag 75% conserved"
23580..23723
/note="MLT1B repeat: matches 1..126 of consensus"
24119..24179

Query Match 88.7%; Score 20.4; DB 9; Length 86719;
 Best Local Similarity 95.5%; Pred. No. 9.7;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 repeat_region /notes="L2 repeat: matches 2688. .2749 of consensus"
 24380. .24429 /notes="25 copies 2 mer to 72% conserved"

QY 1 CAGGAGATCTGAGATTATG 22
 |||||
 Db 11441 CAGGAGATCTGAGATTATG 11462

RESULT 9
 AC079174
 LOCUS
 DEFINITION Homo sapiens 12 BAC RP11-651L5 (Roswell Park Cancer Institute Human
 BAC library) complete sequence.
 AC079174
 AC079174.20 GI:14277179
 HTG.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 87817)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaraturunge,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R.,
 Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
 Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
 Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
 Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
 Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
 Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
 Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsai,F.,
 Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kuresni,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulseghe,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
 Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P.,
 Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
 Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
 Morris,S., Moser,M., Neal,D., Nelson,D., Newtonson,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Oguh,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Savory,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
 Shoohtari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A.,
 Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
 Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
 Thomas,N., Thomas,S., Umani,K., Vasquez,L., Vera,V., Villalon,D.,
 Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
 Washington,C., Watlington,S., Williams,G., Williamson,A.,
 Wlecyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
 Zorilla,S., Kucherlapati,R., Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 87817)
 Worley,K.C.
 Direct Submission
 Submitted (23-AUG-2000) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 87817)
 Worley,K.C.
 Direct Submission
 Submitted (01-JUN-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 87817)
 Worley,K.C.
 Direct Submission
 Submitted (05-JUN-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 5 (bases 1 to 87817)
 Worley,K.C.
 Direct Submission
 Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 6 (bases 1 to 87817)
 Worley,K.C.
 Direct Submission
 Submitted (31-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 7 (bases 1 to 87817)
 Worley,K.C.
 Direct Submission
 Submitted (24-JUN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 8 (bases 1 to 87817)
 Worley,K.C.
 Direct Submission
 Submitted (20-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jun 1, 2001 this sequence version replaced gi:14150295.
 INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
 gc-help@bcm.tmc.edu
 CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.
 ANNOTATION OF FEATURES:
 STSS are identified using ePCR (Genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.
 Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.
 Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.
 SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 clones with no ambiguities or 2 chemistries with a minimum of 2
 reads and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.
 QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found

at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 188340
Phrap values in estimate: 187457
Average error rate (BCM-Phrap estimate): 3.38563e-05
Fraction of Phrap values less than 40 : 0.00584134
Number of consensus changing edits: 44
Number of N's in consensus : 0

----- Consensus changing edits -----
Position Original+Context Edited+Context
576 tcttagtttt(n)agaattaaagg tcttagtttt(a)agaattaaagg
7641 cctctctctt(n)gcccaataca cctctctctt(g)gcccaataca
7719 ctggtgctagt(n)ctgtacattt ctggtgctagt(c)ctgtacattt
7817 tgaagaaca(n)gactgagctg tgaagaaca(a)gactgagctg
19224 gaggacttgc(n)gccgngcng gaggacttgc(c)gccgngcng
19230 ttgngccgc(n)gngcngcng ttgngccgc(c)gngcngcng
19233 cngccgcgc(n)gccgngcng cngccgcgc(c)gccgngcng
19240 ngngccgc(n)gccgngcng ngngccgc(c)gccgngcng
22772 cactaggcca(n)cgcggtggc cactaggcca(g)cgcggtggc
23955 ggaatgagta(n)gtgacattt ggaatgagta(g)gtgacattt
46008 atgacattga(n)atggtcactg atgacattga(g)atggtcactg
54974 ctaactacat(n)cacagctcca ctaactacat(c)cacagctcca
57253 attatata(n)tnnnnnntta attatata(a)ttttattta
57255 ttatatant(n)nnnnnttat ttatatant(t)ttttattt
57256 tatatatnn(n)nnnnnttat tatatatnn(t)ttttattt
57257 atatatnn(n)nnnnnttat atatatnn(t)ttttattt
57258 tatatnnnn(n)nnnnnttat tatatnnnn(a)ttttattt
57259 atatatnn(n)nttatatat atatatnn(t)ttttattt
57260 tatatnnnn(n)tttatatat tatatnnnn(t)ttttattt
57310 tatatttta(n)atattatata tatatttta(t)atattatata
58321 ataattttg(n)atttttaga ataattttg(t)atttttaga
58360 gttgggcagg(n)tggtctcgaa gttgggcagg(c)tggtctcgaa
58403 acctcagct(n)ccaaagtgt acctcagct(c)ccaaagtgt
58412 tnccaagt(n)tggnntaca tnccaagt(c)tggnntaca
58417 aagtgntggg(n)ttacagcat aagtgntggg(a)ttacagcat
58457 aactatttgg(n)attccnaga aactatttgg(t)attccnaga
58464 tggnaattca(n)agacattatt tggnaattca(a)agacattatt
58477 acattattgt(n)tgatctgt acattattgt(t)tgatctgt
75443 taaggaggga(n)acgtagaaat taaggaggga(g)acgtagaaat
79437 aaaaagca(n)ntatagaaa aaaaagca(a)ntatagaaa
79438 aaaaagcan(n)ntatagaaa aaaaagcan(a)ntatagaaa
79440 aaaaagcann(n)atagaaaa aaaaagcann(a)atagaaaa
83879 ttttttttt(n)ncgtnttaa ttttttttt(t)ncgtnttaa
83880 ttttttttn(n)ctgntntaa ttttttttn(t)ctgntntaa
83885 ttttnctgt(n)ntaaatagt ttttnctgt(t)ntaaatagt
83886 tttctgag(n)taaatagtc tttctgag(t)taaatagtc
83887 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83888 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83889 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83890 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83891 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83892 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83893 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83894 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83895 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83896 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83897 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83898 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83899 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83900 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83901 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83902 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83903 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83904 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83905 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83906 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83907 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83908 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83909 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83910 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83911 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83912 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83913 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83914 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83915 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83916 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83917 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83918 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83919 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83920 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83921 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83922 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83923 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83924 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83925 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83926 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83927 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83928 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83929 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83930 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83931 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83932 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83933 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83934 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83935 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83936 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83937 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83938 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83939 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83940 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83941 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83942 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83943 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83944 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83945 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83946 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83947 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83948 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83949 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83950 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83951 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83952 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83953 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83954 tttctgag(n)adggtctcac tttctgag(c)adggtctcac
83955 tttctgag(n)adggtctcac tttctgag(c)adggtctcac

----- Distribution of Quality < 40 Bases -----
1000
900
800
700
Query Match 88.7%; Score 20.4; DB 9; Length 87817;
Best Local Similarity 95.5%; Pred. No. 9.7;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCTCGAGATTATGTG 22
|||||
Db 38408 CAGGAGATCTCGAGATTATGTG 38429
|||||
RESULT 10
HSA132411 88665 bp DNA linear HTG 22-JUN-2001
LOCUS Homo sapiens chromosome 3p21.3, *** SEQUENCING IN PROGRESS ***, 6
DEFINITION unordered pieces.
ACCESSION AJ132411
VERSION AJ132411.1 GI:6599072
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Yang, Y., Kiss, H., Kost-Alimova, M., Kedra, D., Fransson, I.,
Serousi, E., Li, J., Szeles, A., Khododnyuk, I., Imreh, M.P., Fodor, K.,
Hadlaczky, G., Klein, G., Dumaneki, J.P. and Imreh, S.
TITLE A 1-Mb PAC contig spanning the common eliminated region 1 (CER1) in
microcell hybrid-derived SCID tumors
JOURNAL Genomics 62 (2), 147-155 (1999)
MEDLINE 20079145
PUBMED 10610706
REFERENCE 2
AUTHORS Kiss, H., Kedra, D., Yang, Y., Kost-Alimova, M., Kiss, C., O'Brien, K.P.,
Fransson, I., Klein, G., Imreh, S. and Dumaneki, J.P.
TITLE A novel gene containing LIM domains (LIMD1) is located within the
common eliminated region 1 (C1CER1) in 3p21.3
JOURNAL Hum. Genet. 105 (6), 552-559 (1999)
MEDLINE 20112414
PUBMED 10647888
REFERENCE 3
AUTHORS Kedra, D.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-1999) Kedra D., Dept. of Molecular Medicine,
Karolinska Hospital, Building L-8.00, Center of Molecular Medicine
(CMM), S-171 76, SWEDEN
COMMENT This sequence is unfinished. It consists of six contigs. The
contigs are separated by gaps of unknown length. Gaps in the sequence are
represented by a stretch of 200 NNNN. . . . All contigs in this submission are in
their correct order and orientation. 1 22904 contig 1 23105 28324
contig 2 28525 36847 contig 3 37048 40370 contig 4 40571 54712 contig 5
54913
88665 contig 6.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 22904: contig of 22904 bp in length
* 22905 23104: gap of 200 bp
* 23105 28324: contig of 5220 bp in length
* 28325 28524: gap of 200 bp
* 28525 36847: contig of 8323 bp in length
* 36848 37047: gap of 200 bp
* 37048 40370: contig of 3323 bp in length
* 40371 40570: gap of 200 bp
* 40571 54712: contig of 14142 bp in length
* 54713 54912: gap of 200 bp
* 54913 88665: contig of 33753 bp in length.
* Location/Qualifiers
1. .88665
FEATURES
source

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3p21.3"

ORIGIN
Query Match      88.7%; Score 20.4; DB 2; Length 88665;
Best Local Similarity 95.5%; Pred. No. 9.7;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
    |||||
Db 54284 CAGGAGATCCTGAGATTATGTG 54305

RESULT 11
AL157403
LOCUS      111026 bp      DNA      linear      HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 1 clone RP5-1158110 map p22.3-31.2, 13
            unordered pieces.
ACCESSION  AL157403
VERSION     AL157403.3 GI:9796550
KEYWORDS   HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  1
AUTHORS   McIay, K.
TITLE     Direct Submission
JOURNAL   Submitted (08-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequests@sanger.ac.uk
            On Aug 12, 2000 this sequence version replaced gi:9212375.

COMMENT   ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: dj1158110
            ----- Summary Statistics
            Sequencing program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator ABI; 0% of reads
            Chemistry: Dye-terminator Big Dye; 99% of reads
            Consensus quality: 105665 bases at least Q40
            Consensus quality: 107614 bases at least Q30
            Consensus quality: 108618 bases at least Q20
            Insert size: 109826; sum-of-contigs
            Quality coverage: 3.50x in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence.
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 7499: contig of 7499 bp in length
            * 7500: gap of 100 bp
            * 10440: contig of 2841 bp in length
            * 10441: 10540: gap of 100 bp
            * 10541: 17763: contig of 7223 bp in length
            * 17764: 17863: gap of 100 bp
            * 17864: 33875: contig of 16012 bp in length
            * 33876: 33975: gap of 100 bp
            * 33976: 43661: contig of 9686 bp in length
            * 43662: 43761: gap of 100 bp
            * 43762: 52351: contig of 8590 bp in length
            * 52352: 52451: gap of 100 bp
            * 52452: 57434: contig of 4983 bp in length

57435: gap of 100 bp
57535: contig of 7165 bp in length
64700: gap of 100 bp
64800: contig of 2980 bp in length
67779: gap of 100 bp
67879: contig of 7874 bp in length
75753: contig of 100 bp
75754: gap of 100 bp
75854: contig of 8888 bp in length
84741: gap of 100 bp
84842: contig of 16184 bp in length
101025: contig of 100 bp
101026: gap of 100 bp
111026: contig of 9901 bp in length.

FEATURES
            Location/Qualifiers
                1..111026
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosomes="1"
                /map="p22.3-31.2"
                /clone="RP5-1158110"
                /clone_lib="RPCI-5"
                1..7499
                /note="assembly_fragment:00120"
                clone_end:SP6
                vector_side:left
                7600..10440
                /note="assembly_fragment:00467"
                fragment_chain:1
                10541..17763
                /note="assembly_fragment:00136"
                fragment_chain:1
                17864..33875
                /note="assembly_fragment:00746"
                fragment_chain:2
                33976..43661
                /note="assembly_fragment:01078"
                fragment_chain:2
                43762..52351
                /note="assembly_fragment:00050"
                52452..57434
                /note="assembly_fragment:00066"
                57535..64699
                /note="assembly_fragment:00599"
                64800..67779
                /note="assembly_fragment:01030"
                67880..75753
                /note="assembly_fragment:00690"
                fragment_chain:3
                75854..84741
                /note="assembly_fragment:01004"
                fragment_chain:3
                84842..101025
                /note="assembly_fragment:00603"
                fragment_chain:3
                101126..111026
                /note="assembly_fragment:00426"
                fragment_chain:3
                clone_end:T7
                vector_side:right

ORIGIN
Query Match      88.7%; Score 20.4; DB 2; Length 111026;
Best Local Similarity 95.5%; Pred. No. 9.7;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
    |||||
Db 55354 CAGGAGATCCTGAGATTATGTG 55375

RESULT 12
HS74M1/c
LOCUS      HS74M1
            113956 bp      DNA      linear      PRI 27-SEP-2000
```

```

DEFINITION      Human DNA sequence from clone RP1-74M1 on chromosome 1p34.3-36.13
                  Contains the EPHB2 gene for a protein tyrosine kinase. Contains
                  ESTs, GSSs, STSs and a CpG island, complete sequence.
ACCESSION       AL035704
VERSION         GI:6165330
KEYWORDS        HTG; CpG island; EPHB2; tyrosine kinase.
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 113956)
Direct Submission
Submitted (27-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 31, 1999 this sequence version replaced gi:6065897.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone configs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP1-74M1 is from the library RPCI-1 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-74M1.

FEATURES             Location/Qualifiers
     source            1..113956
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosomes="1"
                        /map="p34.3-36.13"
                        /clone="RP1-74M1"
                        /clone_lib="RPCI-1"
                        30..107
repeat_region         /note="MIR repeat: matches 39. .130 of consensus"
repeat_region         527..624
                        /note="MIR repeat: matches 47. .151 of consensus"
repeat_region         804..861
                        /note="MIR repeat: matches 88. .145 of consensus"
repeat_region         880..1007
                        /note="AluJb repeat: matches 4. .125 of consensus"
repeat_region         1008..1294
                        /note="AluX repeat: matches 3. .282 of consensus"
repeat_region         1295..1439
                        /note="AluJb repeat: matches 125. .288 of consensus"
repeat_region         1496..1779
                        /note="AluSg repeat: matches 1. .286 of consensus"
repeat_region         1915..1956
                        /note="L1MB5 repeat: matches 6129. .6172 of consensus"
gene                 2246..35193
                        /gene="EPHB2"
mRNA                 join(<2246..2370,12771..12933,15359..15467,16298..16365,
25874..25996,26597..26844,27840..28055,28909..29058,
30269..30462,32331..32486,33349..33451,33545..33757)
                        /gene="EPHB2"
                        /note="continues in dj61A9 (AL035703), gene dj61A9.1",
                        gene dj61A9.1
match: proteins      Sw:Q90344 Tr:Q43569 Tr:O57458 Sw:P28693
                        Sw:P54763 Sw:P29323 Sw:P54762 Tr:Q91736 Sw:P09759
                        Tr:Q07494 Tr:Q07498 Sw:P54754 Tr:Q91735 Sw:P54753
                        Tr:Q95143 Sw:Q07497 Tr:P21709 Sw:Q05688 Tr:Q9Y130
                        Tr:Q9Y1Y3 Tr:Q9PWR5 Sw:P08069 Tr:Q9V4E5 Sw:P54761
                        Sw:P29317 Tr:Q9XZL6 Tr:Q9UBV3 Tr:Q9VE52 Tr:Q9UBV5

```

```

/gene="EPHB2"
/product="dj74M1.1.1 (tyrosine kinase isoform 1)"
/note="match: CDNAB: Em:AF025304 Em:X75208 Em:AF037331
Em:AF037332 Em:AF037333 Em:U140636 Em:U14164 Em:AF026039
Em:X91737 Em:D31661 Em:L36643 Em:L43620 Em:M62325
Em:L43621 Em:L41939 Em:Z19110 Em:L25890 Em:X76011
Em:D37827 Em:M59814 Em:R09831 Em:Z19061 Em:U11493
Em:U23783 Em:Z49086 Em:U06834 Em:X65138 Em:S57188
Em:U07695 Em:Z19059 Em:AU005029 Em:X13411 Em:D38174
Em:Z49085 Em:D11717 Em:AJ236868 Em:AB025542 Em:M59371
Em:U07634 Em:D14717 Em:AJ005026 Em:X76010 Em:X78339
Em:Z19058 Em:AB025543 Em:U03910 Em:X76012 Em:U89379
Em:M83941 Em:AB025543 Em:M68514 Em:AJ236867 Em:X60380
Em:AB040892 Em:X59290 Em:X15345 Em:L14782 Em:X52822
Em:L14823 Em:L14445 Em:AF041811 Em:L14446 Em:L14447
Em:AF176552 Em:AF216772 Em:AF216773 Em:AF245114
Em:AF216799 Em:AF237766
match: ESTs: Em:AI940384 Em:AW062316 Em:AW351853 Em:M85491
Em:AI940301 Em:AW374328 Em:AA341538 Em:T04965 Em:AW374262
Em:AA298037 Em:AA569391 Em:AI742214 Em:AW662549
Em:AA552367 Em:AI038197 Em:W96473 Em:W96506 Em:AI809403
Em:AA387292 Em:AA505599 Em:AA573757 Em:AI216739
Em:AI739658 Em:W73189 Em:AI445236 Em:AA622626 Em:AI268257
Em:AW162602 Em:AA505740 Em:AA552072 Em:W72792 Em:AI345860
Em:W76439 Em:AI702993 Em:AI476586 Em:AW365103 Em:AA570007
Em:AW168439 Em:AI344182 Em:AI911816 Em:AI345870
Em:AW374261 Em:AA161293 Em:AA524449 Em:AW582672
Em:AA505550 Em:AA505559 Em:AA515992"
/evidence="not experimental"
join(<2246..2370,12771..12933,15359..15467,16298..16365,
25874..25996,26597..26844,27840..28055,28909..29058,
30269..30462,32331..32486,33349..33451,33545..33757)
/gene="EPHB2"
/product="dj74M1.1.2 (tyrosine kinase isoform 2)"
/note="match: CDNAB: Em:L25890 Em:L41939 Em:AF025304
Em:D31661 Em:X91737 Em:M62325 Em:U14164 Em:L40636
Em:L36643 Em:X76011 Em:X75208 Em:AF026039 Em:L43620
Em:AF037333 Em:D37827 Em:AF037332 Em:L43621 Em:Z19110
Em:AF037331 Em:M59814 Em:E09831 Em:Z19061 Em:U23783
Em:Z49086 Em:U11493 Em:E11714 Em:Z19059 Em:X65138
Em:U06834 Em:X13411 Em:D38174 Em:AU005029 Em:Z49085
Em:S57168 Em:U07695 Em:D14717 Em:U07634 Em:AU005026
Em:AB025542 Em:M59371 Em:X76010 Em:X78339 Em:AJ236868
Em:U03910 Em:U89379 Em:M68514 Em:M83941 Em:AU005030
Em:Z19058 Em:X76012 Em:AB025543 Em:AJ236867 Em:X60380
Em:X59290 Em:AB040892 Em:X15345 Em:L14782 Em:L14445
Em:AF041811 Em:L14823 Em:X52822 Em:L14446 Em:L14447
Em:AF176552 Em:AF216772 Em:AF216773 Em:AF245114
Em:AF216799 Em:AF237766
match: ESTs: Em:AI940384 Em:AW062316 Em:AW351853 Em:M85491
Em:AW374328 Em:AA569391 Em:AW374261 Em:AW374262 Em:U04965
Em:AA298037 Em:AI940301 Em:AA341538 Em:AW662549
Em:AA552367 Em:AI038197 Em:W96473 Em:W96506 Em:AI809403
Em:AA505599 Em:AA387292 Em:AA573757 Em:AI216739
Em:AI739658 Em:W73189 Em:AI445236 Em:AA622626 Em:AI268257
Em:AW162602 Em:AA505740 Em:AA552072 Em:W72792 Em:AI345860
Em:W76439 Em:AI702993 Em:AI476586 Em:AW365103 Em:AA570007
Em:AW168439 Em:AI344182 Em:AI911816 Em:AI345870
Em:AA161293 Em:AA524449 Em:AW582672 Em:AI345870
Em:AA505559 Em:AA515992 Em:AI742214"
/evidence="not experimental"
join(<2246..2370,12771..12933,15359..15467,16298..16365,
25874..25996,26597..26844,27840..28055,28909..29058,
30269..30462,32331..32486,33349..33451,33545..33757)
/gene="EPHB2"
/note="continues in dj61A9 (AL035703), gene dj61A9.1",
gene dj61A9.1
match: proteins      Sw:Q90344 Tr:Q43569 Tr:O57458 Sw:P28693
                        Sw:P54763 Sw:P29323 Sw:P54762 Tr:Q91736 Sw:P09759
                        Tr:Q07494 Tr:Q07498 Sw:P54754 Tr:Q91735 Sw:P54753
                        Tr:Q95143 Sw:Q07497 Tr:P21709 Sw:Q05688 Tr:Q9Y130
                        Tr:Q9Y1Y3 Tr:Q9PWR5 Sw:P08069 Tr:Q9V4E5 Sw:P54761
                        Sw:P29317 Tr:Q9XZL6 Tr:Q9UBV3 Tr:Q9VE52 Tr:Q9UBV5

```

CDS

```
Sw:Q03145 Tr:Q91694 Tr:Q91776 Sw:P07948 Tr:Q9PVU9
Tr:Q91845 Tr:P89459 Sw:Q03484 Sw:P08125 Tr:O08644
Tr:O73876 Sw:P29319 Sw:Q02466 Sw:P10936 Sw:P25911
Sw:O15197 Tr:Q62636 Tr:Q952C0 Tr:Q90284 Tr:Q13332
Tr:Q63130 Tr:Q62662 Tr:Q61431 Tr:Q9VUC8 Tr:Q63132
Tr:Q60705 Sw:P41239 Tr:Q90823 Sw:Q60750 Sw:Q03127
Tr:Q9QVY9 Tr:Q9WTL4 Tr:Q93596 Tr:Q93597 Sw:P24062
Tr:Q90601 Tr:Q9XHQ4 Tr:Q92629 Tr:Q905A8 Tr:Q93082
Sw:P08922 Tr:Q73875 Tr:Q91433 Tr:Q42362 Tr:Q07912
Tr:Q15850 Sw:Q07407 Sw:Q04736 Tr:P79726 Tr:Q73878
Tr:Q73798 Tr:Q935Y2 Tr:Q15220 Tr:Q9XE13 Tr:Q13064
Sw:Q02080 Tr:Q13147 Tr:Q9VD94 Sw:Q03137 Tr:Q9PVU0
Sw:P03324 Tr:Q9PVL1 Tr:Q9PVL2 Tr:Q9PVL3 Tr:Q9PVL4
Tr:Q28639 Sw:P32577 Sw:P00529 Tr:Q9VX2 Tr:Q9UHQ6
Sw:P54758 Sw:P00527 Tr:Q24145 Tr:Q43516 Sw:Q93105
Tr:Q64483 Sw:P54760 Tr:Q07701 Tr:Q62214 Sw:P07949
Sw:P14616 Sw:P54764 Tr:Q90478 Tr:P78440 Tr:Q63184
Tr:Q96391 Tr:Q9U7P9 Tr:Q9U8V0 Tr:Q9U8V1 Sw:Q07496
Tr:Q9U8V2 Tr:Q9U531 Tr:Q9U8V1 Tr:Q9U8V2 Tr:Q9V7K5
Tr:Q9U8V3 Tr:Q9U8V6 Tr:Q9U8V9 Tr:Q9U8V12 Tr:Q15250
Sw:Q00495 Sw:P29320 Sw:P00543 Sw:Q06074 Sw:Q09147
Sw:P07332 Sw:P18106 Tr:Q92F58 Tr:Q9U8W2 Tr:P93604
Tr:Q9U8W4 Sw:P29597 Sw:P29400 Sw:P08941 Tr:Q9WEM2
Sw:P15208 Sw:P15127 Tr:Q07438 Tr:Q054951 Tr:Q9VQY1
Sw:Q07014 Sw:P41241 Sw:Q25197 Sw:P41240 Tr:Q9VHE8
Tr:Q9V9V5 Sw:P42689 Tr:Q02742 Tr:Q49202 Sw:Q29000
Sw:P42687 Tr:Q43718 Sw:Q64632 Tr:Q27576 Tr:Q054967
Tr:Q9XDH2 Sw:P42685 Tr:Q93457 Tr:Q9YE47 Tr:Q9V6K3
Tr:Q49205 Tr:Q64434 Sw:P29318 Sw:P09208 Tr:Q13882
Sw:Q01955 Sw:P35546 Sw:P27447 Sw:P09760"
/codon_start=1
/avoidance=not_experimental
/product="dJ74M1.1.2 (tyrosine kinase isoform 2)"
/protein_id="CAC10351.1"
/db_xref="GI:12314011"
/db_xref="GCA:Q9H4H3"
/db_xref="UniProt/TREMBL:Q9H4H3"
/translation="PSAVSIMHQVSTVDSITLSWSQPPQNGVILDYLOYYEKELS
EYNATAIKSTPTVTQGLKAGAVYFVRARTVAGYRGSKMTFQMTAEAYOTS1
QEKPLIGSSAGLVFLIAVVAIVCNRRGFERADSEYTDKLOHTSGHMTGPMK
IYIDPTEDPNEAVEFAKEIDISCVLEQVIGAGEGECVSHLKPGEIEIYFAL
KTLKSGYSEKQRDFSEASIMGQDFPHVHLHGVVTKSTPMIITEFWNGSLDSF
LRQNDQGFVIQVGLMGIAGMGKKKIDFGRGRIQGIFFKEDSHKESDCSCGG"
LEDDSTPTGTSALGKPIRTWPAEAIQYKFTSASDVMSYGVIMVMEVMSYGERP
YWNMTQDVINAIEODYRPLPPMDPCPSALHQLMDCQKDRNRPKFGQIVNTLDKMI
RNPSLKAMAPLSSGILPLLDRTIPDVTSENVDEWLEAIKMGQVKSFAAGFTSF
DVVSOMMEDILVGVTLAGHOKILNSIQVRAQNNQIQSVGGQPLRRPRATGRTK
RCQPRVTKTNCNSGKKGKMGKKIDFGRGRIQGIFFKEDSHKESDCSCGG"
Join(-c2248. .2370,12771. .12933,15359. .15467,16298. .16365,
25874. .25996,26597. .26844,27840. .28055,28909. .29058,
30269. .30462,32331. .32486,33349. .33457)
/genes="EPHB2"
/note="continues in dJ61A9 (AL035703), gene dJ61A9.1
```

```
Query Match 88.7%; Score 20.4; DB 9; Length 113956;
Best Local Similarity 95.5%; Pred. No. 9.7;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 94966 CAGGAGTCTGAGATTATGTG 94945
```

```
RESULT 13
AC106732/c
LOCUS AC106732 119012 bp DNA linear PRI 27-MAR-2002
DEFINITION Homo sapiens chromosome 5 clone CTD-2275D24, complete sequence.
AC106732
ACCESSION AC106732.2 GI:19747160
VERSION AC106732.2
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE 1 (bases 1 to 119012)
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS Direct Submission
TITLE 2 (bases 1 to 119012)
JOURNAL DOE Joint Genome Institute.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (12-JAN-2002) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS 3 (bases 1 to 119012)
JOURNAL DOE Joint Genome Institute.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS 4 (bases 1 to 119012)
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.5.
FEATURES
source
1..119012
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2275D24"
ORIGIN
Query Match 88.7%; Score 20.4; DB 9; Length 119012;
Best Local Similarity 95.5%; Pred. No. 9.7;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 72868 CAGGAGTCTGAGATTATGTG 72847

RESULT 14
AC002416/c
LOCUS AC002416 128915 bp DNA linear PRI 29-JAN-1998
DEFINITION Human Chromosome X, complete sequence.
AC002416
ACCESSION AC002416.1 GI:2822140
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 128915)
TITLE Chen, E., Brownstein, B.H., States, D.J., Schlessinger, D. and
JOURNAL Mazzarella, R.
REFERENCE
AUTHORS Direct Submission
TITLE 2 (bases 1 to 128915)
JOURNAL Brownstein, B.H., States, D.J. and Mazzarella, R.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,
JOURNAL Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA
REFERENCE
AUTHORS 3 (bases 1 to 128915)
JOURNAL Brownstein, B.H., States, D.J. and Mazzarella, R.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (29-JAN-1998) Center for Genetics in Medicine, Box 8232,
JOURNAL Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA
```

COMMENT

On Jan 29, 1998 this sequence version replaced gi:2323255.
 Current status of this project is available at:
 'http://www.ibc.wustl.edu/cgm/seq_projects.html'
 Submitted by:
 Ellison Chen,
 Advanced Center for Genetic Technology,
 Applied Biosystems Division of Perlin Elmer Corp.,
 850 Lincoln Center Drive,
 Foster City, CA 94404 USA
 e-mail: ellison@genseq.apldbio.com

and

Buddy Brownstein,
 Center for Genetics in Medicine,
 Washington University School of Medicine, Box 8232
 4566 Scott Avenue,
 St. Louis, MO 63110, USA
 e-mail: buddy@genetics.wustl.edu

and

David J. States,
 Institute for Biomedical Computing
 Washington University in St. Louis
 700 South Euclid Ave.
 St. Louis, MO 63108 USA
 e-mail: states@ibc.wustl.edu.

FEATURES

source
 1. .128915
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 88.7%; Score 20.4; DB 9; Length 128915;
 Best Local Similarity 95.5%; Pred. No. 9.7;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22

Db 105596 CAGGAGATCCTGAGATTATGTG 105575

RESULT 15

AC068311/c

LOCUS 139203 bp DNA linear HTG 31-MAY-2000
 DEFINITION Homo sapiens chromosome 3 clone RP11-109P19 map 3p, WORKING DRAFT
 SEQUENCE, 47 unordered pieces.

ACCESSION

AC068311.3 GI:8121291

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

Cai, T., Dong, L., Shen, Y., Tan, X., Wang, H., Xi, Y., Xu, Y., Yao, Z.,
 Ma, Q., Rong, L., Zhou, X., Zhou, Y. and Qiang, B.
 Chromosome 3p genomic sequence

TITLE

Unpublished

JOURNAL

REFERENCE

Wang, J., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,
 Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,
 Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,
 Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D.,
 Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J.
 and Yang, H.

TITLE

Direct Submission

Submitted (02-MAY-2000) Human Genomic Center, Institute of

Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing

100101, P.R.China

COMMENT

On May 31, 2000 this sequence version replaced gi:8050886.

-----Genome Center

Center:Beijing Center

Center code:Beijing

Website:http://hgc.igtp.ac.cn

http://www.genomics.org.cn

Contact:hgc@igtp.ac.cn

----- Project Information

Center project name:11% project

Center clone name: RP11-109P19

----- Summary Statistics

Sequencing vector: pUC18; 100% of reads

Chemistry: Dye-terminator; ET 55% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 142729 bases at least Q40

Consensus quality: 157582 bases at least Q30

Consensus quality: 163059 bases at least Q20

Insert size: 105278; sum-of-contigs

Quality coverage: 7.00x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 47 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1213: contig of 1213 bp in length

* 1214 1313: gap of unknown length

* 1314 2442: contig of 1129 bp in length

* 2443 2542: gap of unknown length

* 2543 4207: contig of 1665 bp in length

* 4208 4307: gap of unknown length

* 4308 5648: contig of 1341 bp in length

* 5649 5748: gap of unknown length

* 5749 5968: contig of 1220 bp in length

* 5969 7068: gap of unknown length

* 7069 8123: contig of 1055 bp in length

* 8124 8223: gap of unknown length

* 8224 9393: contig of 1170 bp in length

* 9394 9494: gap of unknown length

* 9494 10687: contig of 1194 bp in length

* 10688 10787: gap of unknown length

* 10788 12037: contig of 1250 bp in length

* 12038 12137: gap of unknown length

* 12138 14344: contig of 2207 bp in length

* 14345 14444: gap of unknown length

* 14445 16007: contig of 1563 bp in length

* 16008 16107: gap of unknown length

* 16108 18196: contig of 2089 bp in length

* 18197 18296: gap of unknown length

* 18297 19846: contig of 1550 bp in length

* 19847 19947: gap of unknown length

* 19947 21757: contig of 1811 bp in length

* 21758 21858: gap of unknown length

* 21858 23938: contig of 2081 bp in length

* 23939 24038: gap of unknown length

* 24039 26931: contig of 2893 bp in length

* 26932 27031: gap of unknown length

* 27032 28895: contig of 1864 bp in length

* 28896 28995: gap of unknown length

* 28996 30602: contig of 1607 bp in length

* 30603 30702: gap of unknown length

* 30703 32033: contig of 1331 bp in length

* 32034 32133: gap of unknown length

* 32134 33637: contig of 1504 bp in length

* 33638 33737: gap of unknown length

* 33738 36656: contig of 2919 bp in length

* 36657 36756: gap of unknown length

* 36757 39152: contig of 2396 bp in length

* 39153 39252: gap of unknown length

| | | | | | |
|---|--------|---------|--------|-------------------|--------------|
| * | 3253 | 40461: | contig | of 1209 | bp in length |
| * | 40462 | 40561: | gap | of unknown length | |
| * | 40562 | 42590: | contig | of 2029 | bp in length |
| * | 42591 | 42690: | gap | of unknown length | |
| * | 42691 | 45244: | contig | of 2554 | bp in length |
| * | 45245 | 45344: | gap | of unknown length | |
| * | 45345 | 47973: | contig | of 2629 | bp in length |
| * | 47974 | 48073: | gap | of unknown length | |
| * | 48074 | 49724: | contig | of 1651 | bp in length |
| * | 49725 | 49824: | gap | of unknown length | |
| * | 49825 | 52090: | contig | of 2266 | bp in length |
| * | 52091 | 52190: | gap | of unknown length | |
| * | 52191 | 54464: | contig | of 2274 | bp in length |
| * | 54465 | 54564: | gap | of unknown length | |
| * | 54565 | 58062: | contig | of 3497 | bp in length |
| * | 58062 | 58161: | gap | of unknown length | |
| * | 58162 | 61037: | contig | of 2876 | bp in length |
| * | 61038 | 61137: | gap | of unknown length | |
| * | 61138 | 63773: | contig | of 2636 | bp in length |
| * | 63774 | 63873: | gap | of unknown length | |
| * | 63874 | 66477: | contig | of 2604 | bp in length |
| * | 66478 | 66577: | gap | of unknown length | |
| * | 66578 | 69285: | contig | of 2708 | bp in length |
| * | 69286 | 71564: | gap | of unknown length | |
| * | 71565 | 71665: | contig | of 2179 | bp in length |
| * | 71666 | 75013: | contig | of 3349 | bp in length |
| * | 75014 | 75113: | gap | of unknown length | |
| * | 75114 | 79238: | contig | of 4125 | bp in length |
| * | 79239 | 79338: | gap | of unknown length | |
| * | 79339 | 82528: | contig | of 3190 | bp in length |
| * | 82529 | 82628: | gap | of unknown length | |
| * | 82629 | 87635: | contig | of 5007 | bp in length |
| * | 87636 | 87735: | gap | of unknown length | |
| * | 87736 | 93180: | contig | of 5445 | bp in length |
| * | 93181 | 93280: | gap | of unknown length | |
| * | 93281 | 98141: | contig | of 4861 | bp in length |
| * | 98142 | 98241: | gap | of unknown length | |
| * | 98242 | 104076: | contig | of 5835 | bp in length |
| * | 104077 | 104176: | gap | of unknown length | |
| * | 104177 | 110939: | contig | of 6763 | bp in length |
| * | 110940 | 111039: | gap | of unknown length | |
| * | 111040 | 115184: | contig | of 4145 | bp in length |
| * | 115185 | 115284: | gap | of unknown length | |
| * | 115285 | 121746: | contig | of 6462 | bp in length |
| * | 121747 | 121846: | gap | of unknown length | |
| * | 121847 | 129746: | contig | of 7900 | bp in length |
| * | 129747 | 129846: | gap | of unknown length | |
| * | 129847 | 139203: | contig | of 9357 | bp in length |

| | |
|--------------|---|
| misc_feature | 1. .1213 |
| misc_feature | /notes/assembly_name:Contig45" 1314. .2442 |
| misc_feature | /note="assembly_name:Contig75" 2543. .4207 |
| misc_feature | /notes/assembly_name:Contig84" 4308. .5648 |
| misc_feature | /notes/assembly_name:Contig86" 5749. .6988 |
| misc_feature | /notes/assembly_name:Contig87" 7069. .8123 |
| misc_feature | /notes/assembly_name:Contig89" 8224. .9393 |
| misc_feature | /note="assembly_name:Contig90" 9494. .10687 |
| misc_feature | /notes/assembly_name:Contig92" 10687. .11813 |

| | | |
|--------------|--------------|---------------------------------|
| misc_feature | 10788..12037 | /note="assembly_name:Contig93" |
| misc_feature | 12138..14344 | /note="assembly_name:Contig94" |
| misc_feature | 14445..16007 | /note="assembly_name:Contig95" |
| misc_feature | 16108..18196 | /note="assembly_name:Contig96" |
| misc_feature | 18297..19846 | /note="assembly_name:Contig97" |
| misc_feature | 19947..21757 | /note="assembly_name:Contig98" |
| misc_feature | 21858..23938 | /note="assembly_name:Contig99" |
| misc_feature | 24039..26931 | /note="assembly_name:Contig100" |
| misc_feature | 27032..28895 | /note="assembly_name:Contig101" |
| misc_feature | 28996..30602 | /note="assembly_name:Contig102" |
| misc_feature | 30703..32033 | /note="assembly_name:Contig103" |
| misc_feature | 32134..33637 | /note="assembly_name:Contig104" |
| misc_feature | 33738..36656 | /note="assembly_name:Contig105" |
| misc_feature | 36757..39152 | |

Query Match 88.7%; Score 20.4; DB 2; Length 139203;
Best Local Similarity 95.5%; Pred. No. 9.7;
Matches 21; Conservative 0; Mismatches 1; Indels 0;

Search completed: August 13, 2005, 05:04:37
Job time : 738.234 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 12, 2005, 22:21:57 ; Search time 184.245 Seconds
(without alignments)
738.985 Million cell updates/sec

Title: US-10-673-854-4

Perfect score: 23

Sequence: 1 caggagatcctgagattatgtgg 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 19.8 | 86.1 | 2995 | 10 | Adf82489 Leukaemia |
| 2 | 18.8 | 81.7 | 1112 | 10 | Adc20846 Human sec |
| 3 | 18.8 | 81.7 | 1112 | 10 | Abt16983 Human sec |
| 4 | 18.8 | 81.7 | 1112 | 10 | Abz67931 Human sec |
| 5 | 18.8 | 81.7 | 1149 | 12 | Adn05709 Antipsori |
| 6 | 18.8 | 81.7 | 1160 | 3 | Aaf22348 Human sec |
| 7 | 18.8 | 81.7 | 1160 | 10 | Adc20223 Human sec |
| 8 | 18.8 | 81.7 | 1160 | 10 | Abt16829 Human sec |
| 9 | 18.8 | 81.7 | 1160 | 10 | Abz67143 Human sec |
| 10 | 18.8 | 81.7 | 1238 | 10 | Adc20845 Human sec |
| 11 | 18.8 | 81.7 | 1238 | 10 | Abt16982 Human sec |
| 12 | 18.8 | 81.7 | 1238 | 10 | Abz67930 Human sec |
| 13 | 18.8 | 81.7 | 1545 | 4 | Aak86702 Human imm |
| 14 | 18.8 | 81.7 | 1545 | 4 | Aak86701 Human imm |
| 15 | 18.8 | 81.7 | 1547 | 4 | Aak86700 Human imm |
| 16 | 18.8 | 81.7 | 1547 | 4 | Aak86700 Human imm |
| 17 | 18.8 | 81.7 | 13919 | 6 | Abx34674 Human mdd |
| 18 | 18.8 | 81.7 | 13919 | 6 | Abk86218 DNA encod |
| 19 | 18.8 | 81.7 | 13919 | 6 | Abk86221 AIP-1/FLA |
| 20 | 18.8 | 81.7 | 21404 | 6 | Abk86220 AIP-1/FLA |
| | | | | | Abk86229 AIP-1/FLA |

| | | | | | | |
|----|------|------|--------|----|------------|--------------------|
| 21 | 18.8 | 81.7 | 55795 | 6 | ABL68242 | Ab168242 Kidney ca |
| 22 | 18.8 | 81.7 | 55795 | 6 | ABL6863 | Ab16863 Kidney ca |
| 23 | 18.8 | 81.7 | 55795 | 6 | ABL68484 | Ab168484 Kidney ca |
| 24 | 18.8 | 81.7 | 55795 | 6 | ABN95045 | Abn95045 Gene #154 |
| 25 | 18.8 | 81.7 | 58320 | 13 | ABD33125 | Abd33125 Human can |
| 26 | 18.8 | 81.7 | 110000 | 10 | ADG70447_0 | Adg70447 Human ANG |
| 27 | 18.8 | 81.7 | 110000 | 10 | ABZ79565_0 | Abz79565 CLLD8 and |
| 28 | 18.8 | 81.7 | 188794 | 12 | ADQ59476 | Adq59476 Human can |
| 29 | 18.8 | 81.7 | 226215 | 11 | ACN45146 | Acn45146 Human gen |
| 30 | 18.2 | 79.1 | 1097 | 12 | ADL90243 | Adl90243 Human gen |
| 31 | 18.2 | 79.1 | 1220 | 12 | ADL90240 | Adl90240 Human enz |
| 32 | 18.2 | 79.1 | 2311 | 12 | ADQ67262 | Adq67262 Novel hum |
| 33 | 18.2 | 79.1 | 2797 | 6 | ABE57465 | Ab157465 Human pro |
| 34 | 18.2 | 79.1 | 3243 | 12 | ADL90241 | Adl90241 Human enz |
| 35 | 18.2 | 79.1 | 3359 | 12 | ADL90242 | Adl90242 Human enz |
| 36 | 18.2 | 79.1 | 3367 | 9 | ARD57344 | Aad57344 Human kin |
| 37 | 18.2 | 79.1 | 138115 | 13 | ABD33312 | Abd33312 Murine ca |
| 38 | 17.8 | 77.4 | 700 | 4 | AH92092 | Aah92092 Human inf |
| 39 | 17.8 | 77.4 | 42024 | 12 | ADQ97373 | Adq97373 Human can |
| 40 | 17.8 | 77.4 | 110000 | 3 | AAF22303_2 | Continuation (3 of |
| 41 | 17.2 | 74.8 | 135 | 5 | ABA16804 | Abal6804 Human ner |
| 42 | 17.2 | 74.8 | 140 | 4 | AAK58947 | Aak58947 Human imm |
| 43 | 17.2 | 74.8 | 148 | 12 | ACH92736 | Ach92736 Human gen |
| 44 | 17.2 | 74.8 | 149 | 4 | AAK58234 | Aak58234 Human imm |
| 45 | 17.2 | 74.8 | 203 | 2 | AAT24266 | Aat24266 Human gen |

ALIGNMENTS

RESULT 1

ADf82489
ID ADF82489 standard; DNA; 2995 BP.
XX
AC ADF82489;
XX
DT 26-FEB-2004 (first entry)
XX
DE Leukaemia-related DNA sequence #3045.
XX
KW Cytostatic; Gene therapy; leukaemia; ss.
XX
OS Unidentified.
XX
PN WO2003039443-A2.
XX
PD 15-MAY-2003.
XX
PF 04-NOV-2002; 2002WO-EF012303.
XX
PR 05-NOV-2001; 2001EP-00126244.
PR 30-APR-2002; 2002EP-00009758.
XX
XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX (UYLU-) UNIV LUDWIG MAXIMILIANS.
XX (HAPE/) HAERLACH T.
XX (SCHG/) SCHOCH C.
XX (KERN/) KERN W.

Haerlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
Bils R, Brors B, Mergenthaler S;
WPI; 2003-505037/47.

Determining the subtype of leukemia cells and whether a patient sample contains leukemia cells or other cells, useful for treating leukemia, comprises determining the expression profile of a group of markers in a patient sample.

Disclosure; SEQ ID NO 3045; 2938pp; English.

The present invention relates to a method (M1) for determining the subtype of leukaemia cells and whether a patient sample contains

CC leukaemia cells. The method comprises determining the expression profile
 CC of a group of markers in a patient sample. The method is useful for
 CC determining the presence of leukaemia cells, its types or subtypes, and
 CC for the preparation of a medicament for treating leukaemia.

XX SQ Sequence 2995 BP; 845 A; 585 C; 700 G; 865 T; 0 U; 0 Other;
 Query Match 86.1%; Score 19.8; DB 10; Length 2995;
 Best Local Similarity 91.3%; Pred. No. 19;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 23
 ||||| ||||| ||||| ||||| |||||
 Db 2029 CAGGAATCTGAGACTATGTG 2051

RESULT 2
 ADC20846
 ID ADC20846 standard; DNA; 1112 BP.
 XX AC ADC20846;
 XX DT 18-DEC-2003 (first entry)
 XX DE Human secreted protein-related DNA sequence #264.
 XX KW gene therapy; human; secreted protein; haemopoietic disorder;
 KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
 KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
 KW leukaemia; wound healing; epithelial cell proliferation disorder;
 KW immune disorder; autoimmune disorder; asthmatic disorder;
 KW cardiovascular disorder; atherosclerosis; myocarditis;
 KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
 KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
 XX OS Homo sapiens.
 XX PN WO200292787-A2.
 XX PD 21-NOV-2002.
 XX PF 26-MAR-2002; 2002WO-US009257.
 XX PR 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX PA Rosen CA, Ruben SM;
 XX PI WPI; 2003-129287/12.
 XX DR
 XX PT New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating hematopoietic or hematologic disorders, e.g.
 PT anemia or hemophilia.
 XX PS Disclosure; SEQ ID NO 800; 1512pp; English.
 XX CC The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for detecting, preventing, diagnosing, prognosticating, treating
 CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
 CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
 CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
 CC wound healing and disorders of epithelial cell proliferation; immune
 CC disorders (e.g. autoimmune disorders and asthmatic disorders);
 CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
 CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
 CC and gastrointestinal disorders (e.g. duodenal ulcers and
 CC gastroenteritis). The present DNA sequence was used in the
 CC exemplification of the invention.

XX SQ Sequence 1112 BP; 267 A; 221 C; 178 G; 446 T; 0 U; 0 Other;
 Query Match 81.7%; Score 18.8; DB 10; Length 1112;
 Best Local Similarity 90.9%; Pred. No. 50;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
 ||||| ||||| ||||| ||||| |||||
 Db 369 CAGGAGATCCTGAGAACATGTG 390

RESULT 3
 ABT16983
 ID ABT16983 standard; DNA; 1112 BP.
 XX AC ABT16983;
 XX DT 03-APR-2003 (first entry)
 XX DE Human secreted protein-related DNA sequence - SEQ ID No 337.
 XX KW Human; gene; ds; protein therapy; immediate hypersensitivity disease;
 KW allergic disorder; asthmatic disorder; gene therapy; secreted protein;
 KW hay fever; allergic conjunctivitis; allergic rhinitis;
 KW binding partner identification; chromosome identification;
 KW radiation hybrid mapping; long-range restriction mapping.
 XX OS Homo sapiens.
 XX PN WO200277188-A2.
 XX PD 03-OCT-2002.
 XX PF 26-MAR-2002; 2002WO-US009239.
 XX PR 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX PA Rosen CA, Ruben SM;
 XX PI WPI; 2003-175010/17.
 XX DR
 XX PT Use of human secreted proteins and nucleic acids for preparing a
 PT diagnostic or pharmaceutical composition for diagnosing or treating
 PT allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic
 PT conjunctivitis or rhinitis.
 XX PS Disclosure; Page 769; 823pp; English.
 XX CC The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for the diagnosis and treatment of allergic disorders, asthmatic
 CC disorders and immediate hypersensitivity diseases (e.g. hay fever,
 CC allergic conjunctivitis and allergic rhinitis). The proteins of the
 CC invention are also useful for identifying a binding partner. The nucleic
 CC acids of the invention are also useful for chromosome identification.
 CC radiation hybrid mapping or long-range restriction mapping. The present
 CC DNA sequence represents a human secreted protein-related DNA sequence

XX SQ Sequence 1112 BP; 267 A; 221 C; 178 G; 446 T; 0 U; 0 Other;
 Query Match 81.7%; Score 18.8; DB 10; Length 1112;
 Best Local Similarity 90.9%; Pred. No. 50;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
 ||||| ||||| ||||| ||||| |||||
 Db 369 CAGGAGATCCTGAGAACATGTG 390

RESULT 4
ABZ67931
ID ABZ67931 standard; DNA; 1112 BP.
XX
XX
AC ABZ67931;
DT 26-MAR-2003 (first entry)
XX
XX
DE Human secreted protein encoding genomic DNA SEQ ID NO 1454.
DE
XX
XX
KW Human; secreted protein; neutropenic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; anti-inflamatory; anti-HIV;
KW vulvar; antibacterial; antiparkinsonian; antiscikling; antianemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; ds.
XX
XX
OS Homo sapiens.
XX
XX
PN WO20027186-A2.
XX
PD 03-OCT-2002.
XX
XX
PF 26-MAR-2002; 2002WO-US009188.
XX
XX
PR 27-MAR-2001; 2001US-027850P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Ruben SM;
XX
XX
DR WPI; 2003-040583/03.
XX
XX
PT New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
XX
PS Disclosure; Page 2064; 2423pp; English.
XX
XX
CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP9470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 1112 BP; 267 A; 221 C; 178 G; 446 T; 0 U; 0 Other;
Query Match 81.7%; Score 18.8; DB 10; Length 1112;
Best Local Similarity 90.9%; Pred. No. 50;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 369 CAGGAGATCCTGAGATTATGTG 390
RESULT 5
ABZ67931
ID ABZ67931 standard; DNA; 1112 BP.
XX
XX
AC ABZ67931;
DT 26-MAR-2003 (first entry)
XX
XX
DE Human secreted protein encoding genomic DNA SEQ ID NO 1454.
DE
XX
XX
KW Human; secreted protein; neutropenic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; anti-inflamatory; anti-HIV;
KW vulvar; antibacterial; antiparkinsonian; antiscikling; antianemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; ds.
XX
XX
OS Homo sapiens.
XX
XX
PN WO20027186-A2.
XX
PD 03-OCT-2002.
XX
XX
PF 26-MAR-2002; 2002WO-US009188.
XX
XX
PR 27-MAR-2001; 2001US-027850P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Ruben SM;
XX
XX
DR WPI; 2003-040583/03.
XX
XX
PT New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
XX
PS Disclosure; Page 2064; 2423pp; English.
XX
XX
CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP9470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 1112 BP; 267 A; 221 C; 178 G; 446 T; 0 U; 0 Other;
Query Match 81.7%; Score 18.8; DB 10; Length 1112;
Best Local Similarity 90.9%; Pred. No. 50;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 369 CAGGAGATCCTGAGATTATGTG 390
RESULT 6
AAF22348
ID AAF22348 standard; cDNA; 1160 BP.
XX
XX
AC AAF22348;
DT 26-MAR-2001 (first entry)
XX
XX
DE Human secreted protein gene 33 SEQ ID NO:43.
XX
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
KW cerebroprotective; neutropenic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulvar; gene therapy; neoplasm;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder; infection;
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW skin aging; food additive; preservative; ss.
XX
XX
OS Homo sapiens.

ADN05709
ID ADN05709 standard; cDNA; 1149 BP.
XX
XX
AC ADN05709;
DT 01-JUL-2004 (first entry)
XX
XX
DE Antipsoriatic cDNA sequence #1084.
XX
XX
KW ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
XX
OS Homo sapiens.
XX
PN WO2004028479-A2.
XX
PD 08-APR-2004.
XX
XX
PF 25-SEP-2003; 2003WO-US030907.
XX
XX
PR 25-SEP-2002; 2002US-0414006P.
XX
XX
PA (GETH) GENENTECH INC.
XX
XX
PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WT;
PI Wu TD;
XX
XX
DR WPI; 2004-305105/28.
DR P-PSDB; ADN05710.
XX
XX
PT New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX
XX
PS Claim 1; SEQ ID NO 2103; 3069pp; English.
XX
XX
CC The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polynucleotides of the invention.
XX
SQ Sequence 1149 BP; 276 A; 219 C; 184 G; 420 T; 0 U; 50 Other;
Query Match 81.7%; Score 18.8; DB 12; Length 1149;
Best Local Similarity 90.9%; Pred. No. 51;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 367 CAGGAGATCCTGAGATTATGTG 388
RESULT 6
AAF22348
ID AAF22348 standard; cDNA; 1160 BP.
XX
XX
AC AAF22348;
DT 26-MAR-2001 (first entry)
XX
XX
DE Human secreted protein gene 33 SEQ ID NO:43.
XX
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
KW cerebroprotective; neutropenic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulvar; gene therapy; neoplasm;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder; infection;
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW skin aging; food additive; preservative; ss.
XX
XX
OS Homo sapiens.

XX WO200061748-A1.
 XX 19-OCT-2000.
 XX 06-APR-2000; 2000WO-US008982.
 XX 09-APR-1999; 99US-0128696P.
 XX 14-JAN-2000; 2000US-0176069P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM, Komatsoulis G;
 XX WPI; 2000-638566/61.
 XX P-PSDB; AAB63081.
 XX New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX
 XX Claim 1; Page 421; 480pp; English.
 XX AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
 CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 CC cerebroprotective; neotropic; neuroprotective; antibacterial; virucide;
 CC fungicide; ophthalmological; and vulnerary. The polynucleotides and
 CC proteins can be used to prevent, treat or ameliorate a medical condition
 CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
 CC sheep. They are also used in diagnosing a pathological condition or
 CC susceptibility to a pathological condition. Disorders which are diagnosed
 CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angioneurosis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
 CC ocular disorders e.g. corneal infection. The polypeptides can also be
 CC used to aid wound healing and epithelial cell proliferation, to prevent
 CC skin aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities, fat content,
 CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
 CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent
 CC sequences used in the exemplification of the present invention
 XX
 XX Sequence 1160 BP; 294 A; 227 C; 184 G; 453 T; 0 U; 2 Other;
 Query Match 81.7%; Score 18.8; DB 3; Length 1160;
 Best Local Similarity 90.9%; Pred. No. 51;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CAGGAGATCCTGAGATTATGTG 22
 Db 370 CAGGAGATCCTGAGACATGTG 391
 RESULT 7
 ADC20223
 ID ADC20223 standard; DNA; 1160 BP.
 XX
 XX ADC20223;
 XX 18-DEC-2003 (first entry)
 DT Human secreted protein coding sequence #162.
 XX gene therapy; human; secreted protein; haemopoietic disorder;
 KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
 KW

KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
 KW leukaemia; wound healing; epithelial cell proliferation disorder;
 KW immune disorder; autoimmune disorder; asthmatic disorder;
 KW cardiovascular disorder; atherosclerosis; myocarditis;
 KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
 KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
 XX Homo sapiens.
 OS WO200292787-A2.
 XX 21-NOV-2002.
 XX 26-MAR-2002; 2002WO-US009257.
 XX 27-MAR-2001; 2001US-0278650P.
 XX 12-SEP-2001; 2001US-00950082.
 XX 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM;
 XX WPI; 2003-129287/12.
 XX New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating hematopoietic or hematologic disorders, e.g.
 PT anemia or hemophilia.
 XX Claim 1; SEQ ID NO 172; 1512pp; English.
 XX The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for detecting, preventing, diagnosing, prognosticating, treating
 CC or ameliorating: hematopoietic or hematological disorders (e.g. anaemia
 CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
 CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
 CC wound healing and disorders of epithelial cell proliferation; immune
 CC disorders (e.g. autoimmune disorders and asthmatic disorders);
 CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
 CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
 CC and gastrointestinal disorders (e.g. duodenal ulcers and
 CC gastroenteritis). The present DNA sequence encodes a human secreted
 CC protein of the invention.
 XX Sequence 1160 BP; 294 A; 227 C; 184 G; 453 T; 0 U; 2 Other;
 SQ
 Query Match 81.7%; Score 18.8; DB 10; Length 1160;
 Best Local Similarity 90.9%; Pred. No. 51;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CAGGAGATCCTGAGATTATGTG 22
 Db 370 CAGGAGATCCTGAGAACATGTG 391
 RESULT 8
 ABT16829
 ID ABT16829 standard; DNA; 1160 BP.
 XX
 XX ABT16829;
 XX 03-APR-2003 (first entry)
 DT Human secreted protein gene sequence - SEQ ID No 78.
 XX Human; gene; ds; protein therapy; immediate hypersensitivity disease;
 KW allergic disorder; asthmatic disorder; gene therapy; secreted protein;
 KW hay fever; allergic conjunctivitis; allergic rhinitis;
 KW binding partner identification; chromosome identification;
 KW radiation hybrid mapping; long-range restriction mapping.
 XX

OS Homo sapiens.
 PN WO2002771188-A2.
 XX
 PD 03-OCT-2002.
 XX
 XX 26-MAR-2002; 2002WO-US0009239.
 PF
 XX 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI
 XX WPI; 2003-175010/17.
 DR
 XX
 XX Use of human secreted proteins and nucleic acids for preparing a
 PT diagnostic or pharmaceutical composition for diagnosing or treating
 PT allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic
 PT conjunctivitis or rhinitis.
 XX
 XX Claim 7; Page 601-602; 823pp; English.
 PS
 XX The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for the diagnosis and treatment of allergic disorders, asthmatic
 CC disorders and immediate hypersensitivity diseases (e.g. hay fever,
 CC allergic conjunctivitis and allergic rhinitis). The proteins of the
 CC invention are also useful for identifying a binding partner. The nucleic
 CC acids of the invention are also useful for chromosome identification,
 CC radiation hybrid mapping or long-range restriction mapping. The present
 CC DNA sequence encodes a human secreted protein of the invention
 XX
 SQ Sequence 1160 BP; 294 A; 227 C; 184 G; 453 T; 0 U; 2 Other;
 Query Match 81.7%; Score 18.8; DB 10; Length 1160;
 Best Local Similarity 90.9%; Pred. No. 51;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CAGGAGATCCTGAGATTATGTG 22
 |||||
 DB 370 CAGGAGATCCTGAGAACATGTG 391
 RESULT 9
 ABZ67143
 ID ABZ67143 standard; cDNA; 1160 BP.
 XX
 AC ABZ67143;
 XX
 DT 26-MAR-2003 (first entry)
 XX
 DE Human secreted protein encoding cDNA SEQ ID NO 263.
 XX
 KW Human; secreted protein; neuroprotective; cytosolic;
 KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW vulnery; antibacterial; antiparkinsonian; antickling; antianaemic;
 KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
 KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
 KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; nephrotropic;
 KW gene therapy; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2002771186-A2.
 XX
 XX 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-US0009188.
 XX

PR 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Ruben SM;
 PI
 XX WPI; 2003-040583/03.
 DR P-PSDB; ABP99722.
 XX
 XX New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
 PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
 PT West Nile fever.
 XX
 XX Claim 7; Page 1308; 2423pp; English.
 PS
 XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections
 XX
 SQ Sequence 1160 BP; 294 A; 227 C; 184 G; 453 T; 0 U; 2 Other;
 Query Match 81.7%; Score 18.8; DB 10; Length 1160;
 Best Local Similarity 90.9%; Pred. No. 51;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CAGGAGATCCTGAGATTATGTG 22
 |||||
 DB 370 CAGGAGATCCTGAGAACATGTG 391
 RESULT 10
 ADC20845
 ID ADC20845 standard; DNA; 1238 BP.
 XX
 AC ADC20845;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human secreted protein-related DNA sequence #263.
 XX
 KW gene therapy; human; secreted protein; haemopoietic disorder;
 KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
 KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
 KW leukaemia; wound healing; epithelial cell proliferation disorder;
 KW immune disorder; autoimmune disorder; asthmatic disorder;
 KW cardiovascular disorder; atherosclerosis; myocarditis;
 KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
 KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200292787-A2.
 XX
 XX 21-NOV-2002.
 XX
 PF 26-MAR-2002; 2002WO-US0009257.
 XX
 PR 27-MAR-2001; 2001US-0278650P.

XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP9470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 1238 BP; 322 A; 231 C; 208 G; 477 T; 0 U; 0 Other;

Query Match 81.7%; Score 18.8; DB 10; Length 1238;
Best Local Similarity 90.9%; Pred. No. 51;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CAGGAGATCCTGAGATATG 22
Db 369 CAGGAGATCCTGAGACATG 390
|||||

RESULT 13

AAK86702/C

ID AAK86702 standard; DNA; 1545 BP.

XX AAK86702;

DT 07-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41514.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 14-AUG-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0234848P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-SEP-2000; 2000US-0236371P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.

XX WO200157182-A2.
PN 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180828P.
PR 04-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234397P.
PR 25-SEP-2000; 2000US-0234398P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR (HUMA-) HUMAN GENOME SCI INC.
PA XX
XX XX

PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 PS Disclosure; SEQ ID NO 41512; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 1547 BP; 576 A; 206 C; 310 G; 455 T; 0 U; 0 Other;

Query Match 81.7%; Score 18.8; DB 4; Length 1547;
 Best Local Similarity 90.9%; Pred. No. 53;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
 |||||
 Db 764 CAGAGGTCCTGAGATTATGTG 743
 |||||

Search completed: August 13, 2005, 04:14:46
 Job time : 189.245 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 03:33:32 ; Search time 56.5213 Seconds
(without alignments)
665.844 Million cell updates/sec

Title: US-10-673-854-4

Perfect score: 23

Sequence: 1 caggagatcctgagattatgtgg 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/prodata/1/ina/5A COMB.seq.*

2: /cgn2_6/prodata/1/ina/5B COMB.seq.*

3: /cgn2_6/prodata/1/ina/6A COMB.seq.*

4: /cgn2_6/prodata/1/ina/6B COMB.seq.*

5: /cgn2_6/prodata/1/ina/PCITUS COMB.seq.*

6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---|
| C 1 | 18.8 | 81.7 | 601 | 4 | US-09-949-016-56405 Sequence 56405, A |
| C 2 | 18.8 | 81.7 | 601 | 4 | US-09-949-016-56406 Sequence 56406, A |
| C 3 | 18.8 | 81.7 | 601 | 4 | US-09-949-016-88904 Sequence 88904, A |
| C 4 | 18.8 | 81.7 | 601 | 4 | US-09-949-016-160361 Sequence 160361, A |
| C 5 | 18.8 | 81.7 | 16505 | 4 | US-09-949-016-11962 Sequence 11962, A |
| C 6 | 18.8 | 81.7 | 16506 | 4 | US-09-949-016-11962 Sequence 11962, A |
| C 7 | 18.8 | 81.7 | 25202 | 4 | US-09-949-016-13151 Sequence 13151, A |
| C 8 | 18.8 | 81.7 | 30053 | 4 | US-09-949-016-16231 Sequence 16231, A |
| C 9 | 18.8 | 81.7 | 47375 | 4 | US-09-949-016-15420 Sequence 15420, A |
| C 10 | 18.8 | 81.7 | 58108 | 4 | US-09-949-016-13383 Sequence 13383, A |
| C 11 | 18.8 | 81.7 | 62386 | 4 | US-09-949-016-12823 Sequence 12823, A |
| C 12 | 18.8 | 81.7 | 62386 | 4 | US-09-949-016-12823 Sequence 12823, A |
| C 13 | 18.8 | 81.7 | 93510 | 4 | US-09-949-016-15095 Sequence 15095, A |
| C 14 | 18.8 | 81.7 | 112623 | 4 | US-09-949-016-14374 Sequence 14374, A |
| C 15 | 18.8 | 81.7 | 162465 | 4 | US-09-949-016-14264 Sequence 14264, A |
| C 16 | 18.8 | 81.7 | 234884 | 4 | US-09-949-016-16420 Sequence 16420, A |
| C 17 | 18.2 | 79.1 | 57761 | 4 | US-09-949-016-13429 Sequence 13429, A |
| C 18 | 17.8 | 77.4 | 601 | 4 | US-09-949-016-134977 Sequence 134977, A |
| C 19 | 17.8 | 77.4 | 24954 | 4 | US-09-949-016-13475 Sequence 13475, A |
| C 20 | 17.8 | 77.4 | 78157 | 4 | US-09-949-016-16466 Sequence 16466, A |
| C 21 | 17.8 | 77.4 | 78157 | 4 | US-09-949-016-16467 Sequence 16467, A |
| C 22 | 17.8 | 77.4 | 256171 | 4 | US-09-949-016-12822 Sequence 12822, A |
| C 23 | 17.8 | 77.4 | 256176 | 4 | US-09-949-016-15524 Sequence 15524, A |
| C 24 | 17.8 | 77.4 | 636591 | 4 | US-09-949-016-11808 Sequence 11808, A |
| C 25 | 17.8 | 77.4 | 636591 | 4 | US-09-949-016-11388 Sequence 11388, A |
| C 26 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-50649 Sequence 50649, A |
| C 27 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-50650 Sequence 50650, A |

| | | | | | |
|------|------|------|------|---|---|
| C 28 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-53774 Sequence 53774, A |
| C 29 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-53775 Sequence 53775, A |
| C 30 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-64795 Sequence 64795, A |
| C 31 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-86530 Sequence 86530, A |
| C 32 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-86531 Sequence 86531, A |
| C 33 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-91958 Sequence 91958, A |
| C 34 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-94411 Sequence 94411, A |
| C 35 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-103437 Sequence 103437, A |
| C 36 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-103438 Sequence 103438, A |
| C 37 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-103439 Sequence 103439, A |
| C 38 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-172197 Sequence 172197, A |
| C 39 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-172198 Sequence 172198, A |
| C 40 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-186421 Sequence 186421, A |
| C 41 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-186422 Sequence 186422, A |
| C 42 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-199070 Sequence 199070, A |
| C 43 | 17.2 | 74.8 | 601 | 4 | US-09-949-016-204673 Sequence 204673, A |
| C 44 | 17.2 | 74.8 | 780 | 4 | US-09-573-080A-178 Sequence 178, App |
| C 45 | 17.2 | 74.8 | 1017 | 4 | US-09-573-080A-177 Sequence 177, App |

ALIGNMENTS

RESULT 1

US-09-949-016-56405/c
; Sequence 56405, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56405
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-56405

Query Match 81.7%; Score 18.8; DB 4; Length 601;
Best Local Similarity 90.9%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
DB 520 CAGGAGATCCTGAGAACATGTG 499

RESULT 2

US-09-949-016-56406/c
; Sequence 56406, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56406
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-56406

Query Match      81.7%; Score 18.8; DB 4; Length 601;
Best Local Similarity 90.9%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
    |||||
DB 528 CAGGAGATCCTGAGACATGTG 507

RESULT 3
US-09-949-016-88904
; Sequence 88904, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 88904
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-88904

Query Match      81.7%; Score 18.8; DB 4; Length 601;
Best Local Similarity 90.9%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
    |||||
DB 85 CAGGAGATCCTGAGACATGTG 106

RESULT 4
US-09-949-016-160361/c
; Sequence 160361, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 160361
; LENGTH: 601

; TYPE: DNA
; ORGANISM: Human
US-09-949-016-160361

Query Match      81.7%; Score 18.8; DB 4; Length 601;
Best Local Similarity 90.9%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
    |||||
DB 878 CAGGAGATCCTGAGACATGTG 899

RESULT 6
US-09-949-016-13836
; Sequence 13836, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13836
; LENGTH: 16506
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13836

Query Match      81.7%; Score 18.8; DB 4; Length 16506;
```

Best Local Similarity 90.9%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 878 CAGGAGATCCTGAGACATGTG 899
|||||

RESULT 7
US-09-949-016-13151
; Sequence 13151, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13151
; LENGTH: 25202
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13151

Query Match 81.7%; Score 18.8; DB 4; Length 25202;
Best Local Similarity 90.9%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 5221 CAGGAGATCCTGAGACATGTG 5242
|||||

RESULT 8
US-09-949-016-16231
; Sequence 16231, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16231
; LENGTH: 30053
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16231

Query Match 81.7%; Score 18.8; DB 4; Length 30053;
Best Local Similarity 90.9%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||

Db 20662 CAGGAGATCCTGAGACATGTG 20683

RESULT 9
US-09-949-016-15420
; Sequence 15420, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15420
; LENGTH: 47375
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(47375)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15420

Query Match 81.7%; Score 18.8; DB 4; Length 47375;
Best Local Similarity 90.9%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 18878 CAGGAGATCCTGAGACATGTG 18899
|||||

RESULT 10
US-09-949-016-13383
; Sequence 13383, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13383
; LENGTH: 58108
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13383

Query Match 81.7%; Score 18.8; DB 4; Length 58108;
Best Local Similarity 90.9%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 33903 CAGGAGATCCTGAGACATGTG 33924
|||||

```
RESULT 11
US-09-949-016-12823
; Sequence 12823, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12823
; LENGTH: 62386
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(62386)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12823

Query Match      81.7%; Score 18.8; DB 4; Length 62386;
Best Local Similarity 90.9%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAGGAGATCCTGAGATTATGTG 22
Db      34283 CAGGAGATCCTGAGACATGTG 34304

RESULT 12
US-09-949-016-12823/c
; Sequence 12823, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12823
; LENGTH: 62386
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(62386)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12823

Query Match      81.7%; Score 18.8; DB 4; Length 62386;
Best Local Similarity 90.9%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAGGAGATCCTGAGATTATGTG 22
Db      34283 CAGGAGATCCTGAGACATGTG 34304

RESULT 13
US-09-949-016-15095
; Sequence 15095, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15095
; LENGTH: 93510
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(93510)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15095

Query Match      81.7%; Score 18.8; DB 4; Length 93510;
Best Local Similarity 90.9%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAGGAGATCCTGAGATTATGTG 22
Db      42219 CAGGAGATCCTGAGAACATGTG 42240

RESULT 14
US-09-949-016-14374/c
; Sequence 14374, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14374
; LENGTH: 112623
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(112623)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14374

Query Match      81.7%; Score 18.8; DB 4; Length 112623;
```


Best Local Similarity 90.9%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
| | | | | | | | | | | | | | | | | | | | | |
Db 75109 CAGGAGATCCTGAGACATGTG 75088

RESULT 15
US-09-949-016-14264/c
; Sequence 14264, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14264
; LENGTH: 162465
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14264

Query Match 81.7%; Score 18.8; DB 4; Length 162465;
Best Local Similarity 90.9%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
| | | | | | | | | | | | | | | | | | | | | |
Db 20162 CAGGAGATCCTGAGACATGTG 20141

Search completed: August 13, 2005, 06:48:56
Job time : 59.5213 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 03:26:33 ; Search time 1478.12 Seconds
(without alignments)
592.293 Million cell updates/sec

Title: US-10-673-854-4
Perfect score: 23
Sequence: 1 caggagatcctgagattatgtgg 23

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_est3:*

4: gb_est4:*

5: gb_est5:*

6: gb_est6:*

7: gb_est7:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 20.4 | 88.7 | 459 | 1 | A1024054 |
| 2 | 20.4 | 88.7 | 605 | 1 | AA719016 |
| 3 | 19.8 | 86.1 | 521 | 2 | BE754118 |
| 4 | 19.4 | 84.3 | 456 | 8 | AQ136698 |
| 5 | 19 | 82.6 | 417 | 9 | CE041276 |
| 6 | 18.8 | 81.7 | 221 | 1 | AA831411 |
| 7 | 18.8 | 81.7 | 257 | 2 | BF806428 |
| 8 | 18.8 | 81.7 | 281 | 8 | AQ069577 |
| 9 | 18.8 | 81.7 | 333 | 6 | BY653205 |
| 10 | 18.8 | 81.7 | 337 | 8 | AQ472951 |
| 11 | 18.8 | 81.7 | 378 | 1 | AA608986 |
| 12 | 18.8 | 81.7 | 386 | 8 | AQ279565 |
| 13 | 18.8 | 81.7 | 401 | 1 | AA569281 |
| 14 | 18.8 | 81.7 | 414 | 1 | AL703856 |
| 15 | 18.8 | 81.7 | 441 | 8 | AQ022522 |
| 16 | 18.8 | 81.7 | 449 | 1 | AA701122 |
| 17 | 18.8 | 81.7 | 467 | 8 | AQ172559 |
| 18 | 18.8 | 81.7 | 470 | 5 | EX101541 |
| 19 | 18.8 | 81.7 | 475 | 8 | AQ471928 |
| 20 | 18.8 | 81.7 | 482 | 8 | B94003 |
| 21 | 18.8 | 81.7 | 487 | 8 | AQ426760 |
| 22 | 18.8 | 81.7 | 488 | 1 | AA708132 |
| 23 | 18.8 | 81.7 | 503 | 8 | AQ734095 |
| 24 | 18.8 | 81.7 | 512 | 8 | AQ204899 |

| | | | | | | | | |
|---|----|------|------|-----|---|----------|-----------|-----------|
| C | 25 | 18.8 | 81.7 | 523 | 8 | AQ544890 | AQ544890 | CITBI-E1- |
| | 26 | 18.8 | 81.7 | 532 | 8 | AQ684672 | AQ684672 | HS 5481_B |
| C | 27 | 18.8 | 81.7 | 564 | 8 | AQ149724 | AQ149724 | HS 3178_A |
| | 28 | 18.8 | 81.7 | 564 | 8 | AQ316239 | AQ316239 | RPCI11-10 |
| | 29 | 18.8 | 81.7 | 565 | 8 | AQ791207 | AQ791207 | HS 4555_B |
| | 30 | 18.8 | 81.7 | 573 | 8 | AQ504120 | AQ504120 | RPCI11-2 |
| | 31 | 18.8 | 81.7 | 592 | 2 | AW964983 | EST377056 | |
| C | 32 | 18.8 | 81.7 | 594 | 8 | AQ478792 | AQ478792 | RPCI11-2 |
| | 33 | 18.8 | 81.7 | 608 | 1 | AI887916 | AI887916 | tk1401.X |
| C | 34 | 18.8 | 81.7 | 614 | 4 | BI061657 | BI061657 | IL3-UT011 |
| | 35 | 18.8 | 81.7 | 628 | 8 | AQ485321 | AQ485321 | RPCI11-2 |
| C | 36 | 18.8 | 81.7 | 629 | 8 | AQ550526 | AQ550526 | RPCI11-4 |
| | 37 | 18.8 | 81.7 | 630 | 4 | BI061656 | BI061656 | IL3-UT011 |
| C | 38 | 18.8 | 81.7 | 637 | 1 | AV684105 | AV684105 | AV684105 |
| | 39 | 18.8 | 81.7 | 638 | 9 | AG160747 | AG160747 | Pan trogl |
| | 40 | 18.8 | 81.7 | 641 | 8 | AQ742110 | AQ742110 | HS 5566_B |
| | 41 | 18.8 | 81.7 | 658 | 9 | AG160363 | AG160363 | Pan trogl |
| | 42 | 18.8 | 81.7 | 659 | 4 | BI062008 | BI062008 | IL3-UT011 |
| | 43 | 18.8 | 81.7 | 663 | 9 | CE227365 | CE227365 | tigr-gss- |
| | 44 | 18.8 | 81.7 | 692 | 5 | BM990062 | BM990062 | UI-H-D10- |
| | 45 | 18.8 | 81.7 | 694 | 7 | CF993883 | CF993883 | AGENCOURT |

ALIGNMENTS

RESULT 1
LOCUS A1024054 459 bp mRNA linear EST 27-AUG-1998
DEFINITION ov72b12.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1642847
3', mRNA sequence.
ACCESSION A1024054
VERSION A1024054.1 GI:3239098
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 459)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-r@mail.nih.gov
CGAP Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
www.biol.llnl.gov/bbrp/image/image.html
Insert length: 1046 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 449.

FEATURES

Location/Qualifiers
1..459
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1642847"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares_testis_NHT"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5']
TGTTACCAATCTGAGTGGAGGCGGCCCAATTTTTTTTTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN
Query Match 88.7%; Score 20.4; DB 1; Length 459;
Best Local Similarity 95.5%; Pred. No. 81;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 15 CAGGAGATCCTGAGATTATGTG 36
|||||

RESULT 2
AA719016 605 bp mRNA linear EST 12-JAN-1999
LOCUS
DEFINITION
AA719016.1 Soares_testis_NHT Homo sapiens cDNA clone 1292523 3',
mRNA sequence.
ACCESSION
AA719016.1 GI:2732115
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 605)
AUTHORS
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1362 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 481.
Location/Qualifiers
1. .605
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="1292523"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares_testis_NHT"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

FEATURES
source

ORIGIN
Query Match 88.7%; Score 20.4; DB 1; Length 605;
Best Local Similarity 95.5%; Pred. No. 83;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 16 CAGGAGATCCTGAGATTATGTG 37
|||||

RESULT 3
BE754118/c
LOCUS
DEFINITION
ACCESSION
BE754118
VERSION
BE754118.1 GI:10168110
KEYWORDS
EST.
SOURCE
Bos taurus (cow)
ORGANISM
Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.

REFERENCE
1 (bases 1 to 521)
AUTHORS
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C.,
Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A.,
Chitko-McKown, C.G., Pettea, G., Holt, I., Karamycheva, S., Liang, F.,
Quackenbush, J. and Keefe, J.W.
TITLE
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL
Genome Res. 11 (4), 626-630 (2001)
MEDLINE
21180013
PUBMED
11282978
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 51 row: E column: 11
Seq primer: ATTGAGTGACACTATAG.
Location/Qualifiers
1. .521
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

ORIGIN
Query Match 86.1%; Score 19.8; DB 2; Length 521;
Best Local Similarity 91.3%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 23
|||||
Db 404 CAGGAGACCTGAGATTCTGTGG 382
|||||

RESULT 4
AQ136698/c
LOCUS
DEFINITION
AQ136698.1 HS_2195_A2_F08_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2195 Col=16 Row=K, genomic survey
sequence.
ACCESSION
AQ136698.1 GI:3523764
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

ORIGIN
Query Match 86.1%; Score 19.8; DB 2; Length 521;
Best Local Similarity 91.3%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 23
|||||
Db 404 CAGGAGACCTGAGATTCTGTGG 382
|||||

ORIGIN
Query Match 86.1%; Score 19.8; DB 2; Length 521;
Best Local Similarity 91.3%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 23
|||||
Db 404 CAGGAGACCTGAGATTCTGTGG 382
|||||

ORIGIN
Query Match 86.1%; Score 19.8; DB 2; Length 521;
Best Local Similarity 91.3%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 23
|||||
Db 404 CAGGAGACCTGAGATTCTGTGG 382
|||||

REFERENCE
 AUTHORS Mahairas,G.C., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 PUBMED 1049764
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2195 row: K column: 16
 Class: BAC ends
 High quality sequence stop: 456.

FEATURES
 source
 1. .456
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=2195 Col=16 Row=K"
 /sex="male"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 84.3%; Score 19.4; DB 8; Length 456;
 Best Local Similarity 90.3%; Pred. No. 2.4e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
 |||||
 Db 456 CAGGAGATCCTGAGANCATGTG 435

RESULT 5
 CE041276 tigr-gss-dog-17000350045460 Dog Library Canis familiaris genomic, GSS 24-SEP-2003
 LOCUS tigr-gss-dog-17000350045460 Dog Library Canis familiaris genomic, genomic survey sequence.
 DEFINITION
 ACCESSION CE041276
 VERSION GSS.
 KEYWORDS
 SOURCE
 ORGANISM Canis familiaris (dog)
 Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (bases 1 to 417)
 Kirchner,B.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Ruch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)
 22875432
 14512627
 Contact: Kirchner BF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirchner@tigr.org
 Class: shotgun.
 Location/Qualifiers
 1. .417
 /organism="Canis familiaris"
 /mol_type="genomic DNA"

FEATURES
 source

ORIGIN

Query Match 82.6%; Score 19; DB 9; Length 417;
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGGAGATCCTGAGATTATG 20
 |||||
 Db 329 AGGAGATCCTGAGATTATG 347

RESULT 6
 AA831411

LOCUS oc60h12.s1 NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:1354151 3'
 DEFINITION similar to contains MER4.t3 MER4 repetitive element ;, mRNA sequence.
 ACCESSION AA831411
 VERSION AA831411.1 GI:2904510
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 221)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 445 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 190.
 Location/Qualifiers
 1. .221
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1354151"
 /issue_type="germlinal center B cell"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP GCBI"
 /note="Vector: p77N3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer
 [5'-TGTACCAATCTGAAGTGGAGCGCCCTCATTTTCTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES
 source

ORIGIN

Query Match 81.7%; Score 18.8; DB 1; Length 221;

Best Local Similarity 90.9%; Pred. No. 4.3e+02; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 62 CAGGAGATCCTGAGATGTG 83

RESULT 7
BF806428 257 bp mRNA linear EST 12-JAN-2001
LOCUS RC2-CI0088-081100-011-f10 CI0088 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF806428
VERSION BF806428.1 GI:12135417
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 257)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC2&t2=RC2-CI0088-081100-011-f10&t3=2000-11-08&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 25

High quality sequence stop: 257.

Location/Qualifiers

FEATURES
source
1..257
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CI0088"
/note="Organ: colon ins; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 257;
Best Local Similarity 90.9%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 233 CAGGAGATCCTGAGATCTGTG 254

RESULT 8

AQ069577/c

LOCUS AQ069577 281 bp DNA linear GSS 04-AUG-1998
DEFINITION HS_2267_B2_G01_MP CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2267 Col=2 Row=N, genomic survey sequence.

ACCESSION AQ069577

VERSION AQ069577.1 GI:3384776

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 281)

AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

PUBMED 10449764

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2267 row: N column: 2

Class: BAC ends

High quality sequence stop: 281.

Location/Qualifiers

FEATURES
source
1..281
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="Plate=2267 Col=2 Row=N"
/sex="male"
/clone_libs="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 81.7%; Score 18.8; DB 8; Length 281;
Best Local Similarity 90.9%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||

Db 194 CAGGAGATCCTGAGAACATGTG 173

RESULT 9

BY653205

LOCUS BY653205 333 bp mRNA linear EST 16-DEC-2002

DEFINITION BY653205 RIKEN full-length enriched, visual cortex Mus musculus cDNA clone K330308P06 3', mRNA sequence.

ACCESSION BY653205

VERSION BY653205.1 GI:27013497

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 333)

AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,

Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrowsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Yang, L., Yuan, Z., Zavalon, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Hayashizaki, Y., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Shingawa, A., and Hayashizaki, Y.

TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL
MEDLINE
PUBMED
22354683
1246851

COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/

Aizawa, K., Akimura, T., Aizawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hiroasawa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

Location/Qualifiers
1. .333
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K530308P06"
/tissue type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match 81.7%; Score 18.8; DB 6; Length 333;
Best Local Similarity 90.9%; Pred. No. 4.5e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 104 CAGGAGATCCTGAGATTATGTG 125
|||||

RESULT 10
LOCUS
AQ472951
DEFINITION
CITBI-El-2589A16.TR CITBI-El Homo sapiens genomic clone 2589A16,
genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AQ472951 337 bp DNA linear GSS 23-APR-1999
CITBI-El-2589A16.TR CITBI-El Homo sapiens genomic clone 2589A16,
genomic survey sequence.
AQ472951
GSS.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 337)
Zhao, S., Adams, M.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and
Venter, J.C.
Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
Map Building
Unpublished (1997)
Other GSSs: CITBI-El-2589A16.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

JOURNAL
COMMENT
Other GSSs: CITBI-El-2589A16.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

Location/Qualifiers
1. .337
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2589A16"
/sex="male"
/cell_type="sperm"
/clone_lib="CITBI-El"
/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"

ORIGIN

Query Match 81.7%; Score 18.8; DB 8; Length 337;
Best Local Similarity 90.9%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 201 CAGGAGATCCTGAGATTATGTG 222
|||||

FEATURES

Location/Qualifiers
1. .333
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2589A16"
/sex="male"
/cell_type="sperm"
/clone_lib="CITBI-El"
/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"

ORIGIN

Query Match 81.7%; Score 18.8; DB 8; Length 337;
Best Local Similarity 90.9%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 201 CAGGAGATCCTGAGATTATGTG 222
|||||

RESULT 11
LOCUS
AQ608986
DEFINITION
af05d07.s1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1030765
3' similar to gb:J05158 CARBOXYPEPTIDASE N 83 KD CHAIN
(HUMAN); contains MER4.b2 MER4 MER4 repetitive element i, mRNA
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AQ608986 378 bp mRNA linear EST 02-MAR-1998
af05d07.s1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1030765
3' similar to gb:J05158 CARBOXYPEPTIDASE N 83 KD CHAIN
(HUMAN); contains MER4.b2 MER4 MER4 repetitive element i, mRNA
sequence.
AA608986
AA608986.1 GI:2457414
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 378)

REFERENCE
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST Project

JOURNAL Unpublished (1997)

COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 533 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 367.

FEATURES
source
1..378
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1030765"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; let strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'] . Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia) , digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN
Query Match 81.7%; Score 18.8; DB 1; Length 378;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 69 CAGGAGATCCTGAGAACATGTG 90

RESULT 12
AQ279565/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AQ279565 386 bp DNA linear GSS 22-NOV-1998
CITBI-E1-2523F6, TR CITBI-E1 Homo sapiens genomic clone 2523F6, genomic survey sequence.

AQ279565
AQ279565.1 GI:3905469
GSS.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 386)
Berty,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready Map Building
Unpublished (1998)
Other GSSs: CITBI-E1-2523F6.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mddams@tigr.org
Clones are available from Research Genetics (info@resgen.com) . BAC end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
source
1..386
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2523F6"
/sex="male"
/cell_type="sperm"
/clone_lib="CITBI-E1"
/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"

ORIGIN
Query Match 81.7%; Score 18.8; DB 8; Length 386;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||
Db 155 CAGGAGATCCTGAGAACATGTG 134

RESULT 13
AA569281
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA569281 401 bp mRNA linear EST 09-SEP-1997
nm31e01.s1 NCI CGAP Lip2 Homo sapiens cDNA clone IMAGE:1061784 similar to contains "Alu repetitive element; contains MER4.t3 MER4 repetitive element ;, mRNA sequence.

AA569281
AA569281.1 GI:2342335
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 401)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 684 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 359.

FEATURES
source
1..401
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1061784"
/tissue_type="liposarcoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Lip2"
/note="Vector: pAMP10; mRNA made from liposarcoma, cDNA made by oligo-dT priming. Non-directionally cloned.

Size-selected on agarose gel, average insert size 600 bp.
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."

ORIGIN

Query Match 81.7%; Score 18.8; DB 1; Length 401;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||

Db 34 CAGGAGATCCTGAGAAATGTG 55
|||||

RESULT 14

AL703856

LOCUS

DEFINITION DKFZp686N0727_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
414 bp mRNA linear EST 04-SEP-2003

ACCESSION

VERSION DKFZp686N0727 5', mRNA sequence.

KEYWORDS

AL703856

SOURCE

EST.

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 414)

Ansorge, W., Wirkner, U., Mewes, H.W., Weil, B. and Wiemann, S.)

EST (Ansorge, W., Wirkner, U., Mewes, H.W., Weil, B. and Wiemann, S.)

Unpublished (1999)

Contact: MIPS

MIPS

Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by EMBL (European Molecular Biology Laboratories,

Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No sl sequence available.

This clone (DKFZp686N0727) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1..414

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp686N0727"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="686 (synonym: hlcc3)"

/notes="Vector: pTriplex2; Site_1: SfIIA; Site_2: SfiIB;

cDNA-collection"

ORIGIN

Query Match 81.7%; Score 18.8; DB 1; Length 414;

Best Local Similarity 90.9%; Pred. No. 4.6e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||

Db 287 CAGGAGATCCTGAGACATGTG 308
|||||

RESULT 15

AQ022522/c

LOCUS

DEFINITION HS_2180_A2_D01_MF CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=2180 Col=2 Row=G, genomic survey

sequence.

ACCESSION AQ022522

VERSION AQ022522.1 GI:3220730

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

10449764

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2180 row: G column: 2

Class: BAC ends

High quality sequence stop: 441.

Location/Qualifiers

1..441

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="Plate=2180 Col=2 Row=G"

/sex="male"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in

E-Coli DH10B"

ORIGIN

Query Match 81.7%; Score 18.8; DB 8; Length 441;

Best Local Similarity 90.9%; Pred. No. 4.7e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGAGATCCTGAGATTATGTG 22
|||||

Db 418 CAGGAGATCCTGAGACATGTG 397
|||||

Search completed: August 13, 2005, 06:45:12
Job time : 1485.12 secs

THIS PAGE BLANK (USPTO)